



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Cíntia Lionela Ambrósio de Menezes

Caracterização estrutural e funcional de uma alfa-N-
arabinofuranosidase recombinante expressa em *Escherichia coli*.

São José do Rio Preto

2019

Cíntia Lionela Ambrósio de Menezes

Caracterização estrutural e funcional de uma alfa-N-
arabinofuranosidase recombinante expressa em *Escherichia coli*.

Dissertação apresentada como requisito para obtenção do título de Mestre em Microbiologia, junto ao Programa de Pós-Graduação em Microbiologia, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador: Prof. Dr. Gustavo O. Bonilla Rodriguez

Coorientadores: Prof^a. Dr^a. Eleni Gomes e Dr. Luciano Takeshi Kishi

São José do Rio Preto

2019

M543c Menezes, Cintia Lionela Ambrósio de
Caracterização estrutural e funcional de uma alfa-N-arabinofuranosidase recombinante expressa em Escherichia coli. / Cintia Lionela Ambrósio de Menezes. -- São José do Rio Preto, 2019
89 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Gustavo Orlando Bonilla Rodriguez

1. Microbiologia. 2. Enzimas bacterianas. 3. Arabinofuranosidase. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Caracterização estrutural e funcional de uma alfa-N-
arabinofuranosidase recombinante expressa em *Escherichia coli*.

Dissertação apresentada como requisito para
obtenção do título de Mestre em Microbiologia, junto
ao Programa de Pós-Graduação em Microbiologia,
do Instituto de Biociências, Letras e Ciências Exatas
da Universidade Estadual Paulista “Júlio de
Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof. Dr. Gustavo Orlando Bonilla Rodriguez
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof(a). Dr(a). Cintia Bittar Oliva
UNESP – Câmpus de São José do Rio Preto

Prof(a). Dr(a). Gabriela Salvador de Amo
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)

São José do Rio Preto

01 de julho de 2019

DEDICATÓRIA

Dedico a Deus!

Aos meus familiares, especialmente meus pais, José e Leonice ao meu irmão, Júlio e ao meu companheiro de vida Adriano, pelo apoio, paciência e por todo o amor;

A vocês, dedico meu trabalho e o meu amor.

AGRADECIMENTOS

Agradeço primeiramente ao Prof. Dr. Gustavo Orlando Bonilla Rodriguez, pela orientação durante esses 2 anos de trabalho, por toda ajuda e confiança.

Ao Prof. Dr. Luciano Takeshi Kishi pela coorientação, pelo conhecimento compartilhado, pela atenção e esclarecimentos.

A Profa. Dra. Eleni Gomes pela coorientação, pela ideia inicial do trabalho e disponibilização do laboratório no qual realizei grande parte do projeto.

As queridas amigas Maira Rodrigues e Juliana Prado, pela companhia, pela amizade, carinho, cumplicidade, sentimentos compartilhados nesses anos de convivência.

As meninas do Laboratório de Bioquímica de Microrganismos e Plantas (LBMP) da UNESP/FCAV de Jaboticabal, pela amizade, pelas dúvidas esclarecidas, pela ajuda nesse trabalho e em todas as atividades no laboratório.

A todos do Laboratório Bioquímica e Microbiologia da UNESP/IBILCE, pela amizade e ajuda nesse trabalho e nas atividades no laboratório, especialmente agradeço ao Roni, Gabriela Salvador, Eduardo, Carlos, Lorena e Jéssica obrigada de coração pela ajuda, foi fundamental para mim.

Aos meus amados pais, José e Leonice, e irmão Júlio Cezar pelo amor, paciência, carinho, cuidado, amparo, ajuda, proteção, pela vida, por tudo! Por serem os melhores espelhos. Tudo o que sou e tudo o que sei devo a eles.

Ao meu amor Adriano Almeida, pelo apoio, pela paciência, pelo carinho, pelo amor, por ter trazido paz ao meu coração e leveza aos dias difíceis.

Ao programa de pós-graduação em Microbiologia e a UNESP.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 à qual agradeço.

A todos que direta ou indiretamente contribuíram para a realização desse trabalho.

Meus sinceros agradecimentos.

RESUMO

As abordagens metagenômicas têm sido amplamente utilizadas para isolar novos biocatalisadores de amostras ambientais, uma vez que as comunidades microbianas podem apresentar grande diversidade e, na grande maioria dos casos, não podem ser cultivadas em laboratório. Dessa maneira, estudos direcionados apenas para o isolamento e cultivo não podem fornecer informações suficientes do papel biológico de uma grande quantidade de microrganismos membros de uma comunidade, sendo estas informações melhor entendidas por abordagem metagenômica. Inicialmente investigamos as características das ORFs presentes no genoma de bactéria do gênero *Chitinophaga* com o intuito de avaliar o potencial biotecnológico da comunidade microbiana, e a partir disso a sequência CB10.2_116 foi selecionada. A clonagem e expressão da α -L-arabinofuranosidase codificada pela sequência CB10.2_116 foi realizada, sendo essa selecionada de um banco de dados (CB10) (<http://200.145.102.111/cb10/>) construído a partir do sequenciamento de um consórcio microbiano. Após ser prospectada por similaridade e analisada *in silico*, a proteína codificada apresenta 780 resíduos de aminoácidos, sendo identificada como uma α -L-arabinofuranosidase (EC 3.2.1.55) da família de glicosil hidrolases 127, com 80,3% de similaridade com uma proteína não caracterizada da bactéria *Chitinophaga jiangningensis*. A clonagem se deu em vetor pET28-a. Para expressar a proteína de interesse, a qual apresenta uma massa molecular estimada de 87,99 kDa em célula competente de *E. coli*, uma colônia transformada foi isolada e inoculada em meio Lúria Bertani (LB) contendo canamicina a 50 μ g/mL, e deixada sob agitação constante a 37°C até atingir a fase logarítmica de crescimento, com DO_{600nm} entre 0,6 e 0,8. A cultura foi induzida com isopropil- β -D- tiogalactopiranosídeo (IPTG). A confirmação da expressão da proteína foi realizada através da eletroforese desnaturante em gel de poliacrilamida 12% - SDS-PAGE. Neste estudo não foi possível alcançar a expressão ideal da enzima de interesse, acredita-se que em virtude do sistema de expressão escolhido.

Palavras-chave: α -L-arabinofuranosidase, enzima recombinante, metagenômica

ABSTRACT

Metagenomic approaches have been widely used to isolate new biocatalysts from environmental samples, since microbial communities can exhibit great diversity and, in the vast majority of cases, can not be grown in the laboratory. Thus, studies directed only at isolation and cultivation can not provide sufficient information on the biological role of a large number of microorganisms members of a community, and this information is better understood by metagenomic approach. We initially investigated the characteristics of the ORFs present in the genome of Chitinophaga bacteria in order to evaluate the biotechnological potential of the microbial community, and from that the sequence CB10.2_116 was selected. The cloning and expression of the α -L-arabinofuranosidase encoded by the sequence CB10.2_116 was performed, being selected from a database (CB10) (<http://200.145.102.111/cb10/>) constructed from the sequencing of a consortium microbial. After being prospected for similarity and analyzed in silico, the encoded protein shows 780 amino acid residues, being identified as an α -L-arabinofuranosidase (EC 3.2.1.55) from the family of glycosyl hydrolases 127, with 80.3% similarity to a uncharacterized protein from the bacterium Chitinophaga jiangningensis. Cloning took place in vector pET28-a. To express the protein of interest, which has an estimated molecular mass of 87.99 kDa in a competent E. coli cell, a transformed colony was isolated and inoculated into Bertani lutein medium (LB) containing 50 μ g / ml kanamycin, and left under constant stirring at 37 ° C until reaching the logarithmic growth stage, with DO600nm between 0.6 and 0.8. The culture was induced with isopropyl- β -D-thiogalactopyranoside (IPTG). Confirmation of protein expression was performed by denaturing electrophoresis on 12% polyacrylamide gel - SDS-PAGE. In this study it was not possible to achieve the ideal expression of the enzyme of interest, it is believed that by virtue of the chosen expression system.

Keywords: α -L-arabinofuranosidase, recombinant enzyme, metagenomics

LISTA DE ABREVIATURAS

AAs	Enzimas auxiliares
ABFs	Arabinofuranosidases
α -L-AFase	α -L-Arabinofuranosidase
CAZymes	Enzimas ativas em carboidratos
CEs	Carboidrato esterases
DNA	Ácido Desoxirribonucléico
DUF	<i>Domain of unknown functions</i>
EC	<i>Enzyme Commission</i>
GH	Glicosil Hidrolase
GTs	Glicosil Transferase
HMMs	<i>Hidden Markov Model</i>
IPTG	Isopropil β -D-tiogalactopiranosídeo
MSA	Alinhamento de múltiplas sequências
NGS	Next-Generation Sequence
DO	Densidade Óptica
ORF	<i>Open Reading Frame</i>
PB	Pares de base
PCR	Reação em Cadeia da Polimerase
PLs	Polissacarídeo liases
pI	Ponto Isoelétrico
SBS	Sequenciamento por síntese
SDS	Dodecil Sulfato de Sódio
XOS	Xilo-oligossacarídeos
WGS	<i>Whole genome shotgun</i>

SUMÁRIO

Capítulo I: Análise <i>in silico</i> do genoma de <i>Chitinophaga</i> sp., e seleção da sequência que codifica uma arabinofuranosidase.....	10
1. INTRODUÇÃO	10
2. REVISÃO DE LITERATURA.....	12
2.1 Sequenciamento de DNA	14
2.2 Uso de ferramentas para montagem de genomas/metagenomas	15
2.3 Estudo de amostras ambientais	17
3. OBJETIVOS	22
4. MATERIAL E MÉTODOS.....	22
4.1 Extração de DNA total e sequenciamento da comunidade bacteriana CB10.	23
4.2 Montagem, classificação taxonômica e perfil funcional da comunidade bacteriana (CB10).	23
4.3 Avaliação do potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica.....	24
4.4 Alinhamento de múltiplas sequências.....	25
4.5 Software de Análise Genética Evolutiva Molecular	25
4.6 Construção de Árvores filogenéticas	26
5. RESULTADOS E DISCUSSÃO	26
5.1 Caracterização da ORF selecionada CB10_2.116 que codifica uma arabinofuranosidase.	33
5.2 Expressão de genes responsáveis pela produção de enzimas degradadoras de lignocelulose.....	36
6. CONCLUSÕES PARCIAIS	39
REFERÊNCIAS	40
Capítulo II: Clonagem e expressão da sequência CB10_2.116.	44
1. INTRODUÇÃO	44
2. REVISÃO DA LITERATURA	46
2.1 Biomassa lignocelulósica.....	46
2.2 Hemicelulose	49
2.3 Famílias de glicosil hidrolases	50
2.4 Aplicações biotecnológicas de arabinofuranosidases	51
2.5 Expressão heteróloga de proteínas em <i>Escherichia coli</i>	54
2.6 Análises <i>In silico</i>	57

3. OBJETIVOS	60
4. MATERIAL E MÉTODOS	61
4.1 Análise <i>in silico</i> das sequências	61
4.2 Clonagem da sequência codificadora de arabinofuranosidase	61
4.3 Eletroforese preparativa	63
4.4 Digestão do vetor e inserto com BamHI	63
4.4.1 Digestão de Vetor	63
4.4.2 Digestão do Inserto.....	63
4.5 Digestão do vetor e inserto com EcoRI	64
4.5.1 Digestão de Vetor	64
4.5.2 Digestão de Inserto.....	64
4.6 Reação de Ligação do fragmento ao vetor de expressão.....	64
4.7 Transformação bacteriana das células de <i>Escherichia coli</i>	66
4.8 Análise de sequenciamento da ORF	67
4.9 Construção do plasmídeo de expressão em <i>E. coli</i> (Fastbio)	67
4.10 Expressão da proteína recombinante	68
4.11 Eletroforese em gel de poliacrilamida (SDS-PAGE)	69
4.12 Recuperação da proteína em fração insolúvel.....	69
4.13 Purificação por cromatografia de afinidade ao metal	70
4.14 Ensaio de atividade enzimática	70
5. RESULTADOS E DISCUSSÃO	70
5.1 Análises <i>in silico</i>	70
5.2 Amplificação e clonagem dos genes no vetor pET28a	76
5.3 Ensaio de expressão e extração das proteínas recombinantes	79
5.4 Purificação e atividade enzimática.....	83
6. CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS	84

Capítulo I: Análise *in silico* do genoma de *Chitinophaga* sp., e seleção da sequência que codifica uma arabinofuranosidase

1. INTRODUÇÃO

As enzimas são biocatalisadores poderosos que possuem a capacidade de aumentar significativamente as taxas de um grande número de reações biológicas e químicas. Atualmente há uma alta demanda pelas enzimas, que são consideradas alternativas mais ecológicas para a síntese química, pois ainda hoje são usados solventes orgânicos prejudiciais para o meio ambiente e elevado consumo de energia. Enzimas microbianas não possuem qualquer natureza perigosa para o ambiente, contribuindo como soluções ecologicamente adequadas na síntese de produtos químicos. O estudo com comunidades microbianas inexploradas para a identificação e isolamento de enzimas é uma área ativa de pesquisa em todo o mundo (MADHAVAN et al., 2017).

Os microrganismos procarióticos possuem uma notável diversidade fisiológica, metabólica e funcional e constituem a fonte mais rica de diversidade genética. O método clássico para prospectar essa informação diversa de genes é cultivando o microrganismo e fazendo a subsequente triagem para o fenótipo desejado. No entanto, cerca de 99,9% dos microrganismos em nichos ambientais não podem ser cultivados por técnicas laboratoriais padrão. Com isso, ferramentas e tecnologias que permitam pesquisas de microbiomas têm crescido em ritmo acelerado graças ao aumento no uso de sequenciadores rápidos e de técnicas ômicas (MADHAVAN et al., 2017 e NAVAS-MOLINA et al., 2017).

É grande a quantidade de dados coletados pelos pesquisadores, o que traz para a comunidade científica desafios no que diz respeito ao armazenamento e análise de dados. Porém, essa riqueza de informações também facilitará a compreensão de mecanismos e interações comunitárias bacterianas como nunca antes, nos conduzindo ao desenvolvimento inovador em diversas áreas, não apenas na saúde humana, mas também na agricultura e biocombustíveis, entre outras aplicações. Um dos maiores desafios enfrentados por pesquisadores que

visam atingir esses objetivos é a capacidade de integrar e correlacionar as enormes quantidades de dados produzidos por essas novas tecnologias e identificar informações biologicamente relevantes que podem ser usadas para formular hipóteses testáveis (NAVAS-MOLINA et al., 2017).

Os esforços de sequenciamento nos últimos anos levaram o número para mais perto de 1.000 filós de bactérias e de arqueas reconhecidos. Toda essa abundância de microrganismos fornece contribuições ao ecossistema que são cruciais para a sustentabilidade local e global. A microbiota dentro e sobre as culturas, árvores e outras plantas, nos solos em que crescem, fornecem nitrogênio, fósforo e outros nutrientes essenciais, além de atuarem na decomposição de poluentes e suprimirem a atividade de microrganismos patogênicos. Reconhecendo o poder inexplorado dos microbiomas do solo e das plantas no aumento da produtividade agrícola, empresas estão investindo milhões de dólares em pesquisa e desenvolvimento nessa área (DUBILIER; MCFALL-NGAI; ZHAO, 2015).

As abordagens metagenômicas têm sido amplamente utilizadas para isolar novos biocatalisadores de amostras ambientais. As comunidades microbianas podem apresentar grande diversidade, e que na grande maioria dos casos não podem ser identificadas em laboratório, pois os estudos direcionados apenas no isolamento e cultivo não podem fornecer informação suficientes do papel biológico de uma grande quantidade de microrganismos membros de uma comunidade, sendo que tais informações podem ser melhor entendidas por abordagem metagenômica (GONG et al., 2012).

Neste estudo investigamos as características das ORF presentes no genoma de bactéria do gênero *Chitinophaga* com o intuito de avaliar o potencial biotecnológico da região CB10.2116 proveniente de estudos metagenômicos. Estudos posteriores puderam identificar que tal sequência pode expressar uma enzima da família de glicosil hidrolase com a importante função de atuar na desconstrução de material lignocelulósico.

2. REVISÃO DE LITERATURA

Em 1985 um avanço experimental mudou radicalmente a maneira como visualizamos o mundo microbiano; Pace e seus colegas criaram um novo ramo da ecologia microbiana utilizando a análise direta das sequências dos genes 5S e 16S rRNA no ambiente para descrever a diversidade de microrganismos em uma amostra ambiental sem o seu cultivo. Os primeiros estudos eram tecnicamente desafiadores, contando com sequenciamento direto de RNA ou sequenciamento de cópias de DNA geradas por transcrição reversa. O próximo avanço técnico chegou com o desenvolvimento da tecnologia de PCR e o desenho de *primers* que podem ser usados para amplificar quase todos os genes. Isso acelerou a descoberta de diversos táxons à medida que os habitats da Terra foram pesquisados pela nova técnica. A aplicação da tecnologia de PCR forneceu uma visão da diversidade microbiana que não foi distorcida pelo viés da cultura e revelou que a maioria não cultivável é altamente diversificada (HANDELSMAN, 2004).

A ciência está apenas começando a entender a importância do mundo microbiano, graças ao desenvolvimento de técnicas como sequenciamento de alto rendimento e baixo custo, avanços na preparação de amostras que permitem aos pesquisadores sequenciar genomas de células individuais, bem como de comunidades microbianas, melhorias no poder de computação e tecnologias de imagem e desenvolvimento de ferramentas de bioinformática importante na compreensão dos dados. Assim, os pesquisadores podem identificar e compreender a função dos microrganismos que não podem ser cultivados em laboratório ou seja, a grande maioria do microbioma terrestre (DUBILIER, MCFALL-NGAI E ZHAO, 2015).

Em virtude do desenvolvimento das tecnologias “ômicas”, como a genômica, transcriptômica, proteômica, metagenômica, o metatranscriptômica e metabolômica, ocorreram avanços importantes principalmente nos campos da saúde, biotecnologia, ecologia e alimentação. Muitos deles se utilizando do estudo com microrganismos e o ambiente ecológico no qual se desenvolvem. Diante disso se pode destacar feitos importantes, como por exemplo: 1) descrever os genomas de uma grande variedade de organismos; 2) estudar a expressão diferencial de genes pelos efeitos do ambiente celular; 3) estudar os efeitos de vários compostos

químicos na síntese de proteínas; 4) estudar a grande variedade de rotas metabólicas, bem como identificar novas vias, entre outros (QUIROGA, 2016).

O termo proteoma definido por Marc Wilkins em 1994, é uma imagem dinâmica de todas as proteínas expressas por um organismo, célula ou compartimento subcelular específico, em um dado tempo e sob certas condições, que constitui a expressão da proteína em uma célula, tecidos ou um determinado organismo. Porém algumas questões complexas devem ser consideradas no entendimento do estudo com expressão de proteínas; um fator a ser considerado diz respeito às modificações que poderão ocorrer na estrutura ou na sequência de proteína de núcleo e modificações pós-tradução, que servem para modificar ou modular a atividade, função ou localização de uma proteína em diferentes contextos fisiológicos ou metabólicos. Finalmente, outro fator de complexidade se refere à multifuncionalidade, em que a mesma proteína pode desempenhar várias funções (BARBOSA et al., 2012; QUERO et al., 2016).

Essa nova abordagem para o estudo de proteínas se desenvolveu graças à integração de tecnologias tais como os métodos de separação de proteínas de alto desempenho, a espectrometria de massa e, acima de tudo, ferramentas em bioinformática. Graças à sua aplicabilidade, a proteômica permite não apenas identificar proteínas, mas também categorizá-las, classificá-las em termos de suas funções e interações. O campo de estudo da proteômica é amplo, e pode ser dividido nas seguintes categorias: a) expressão proteômica, que visa identificar as proteínas presentes na amostra e obter todas as informações sobre abundâncias, modificações pós traducionais e localização subcelular; b) proteômica estrutural, o que permite obter a estrutura tridimensional das proteínas, e c) proteômica funcional, cuja finalidade é determinar a função de proteínas e compreender as interações que podem ser executadas (QUERO et al., 2016).

2.1 Sequenciamento de DNA

A sequência das bases nucleotídicas nas cadeias dos ácidos nucleicos contém a informação para as propriedades hereditárias e bioquímicas da vida terrestre. Portanto, a capacidade de medir ou inferir tais sequências é necessária para a pesquisa biológica. A história do sequenciamento genômico teve início em meados da década de 70, primeiramente com a elaboração da metodologia descrita por Frederick Sanger e colaboradores em 1975, a qual ficou conhecida por método de Sanger. De forma independente, Maxam e Gilbert desenvolveram, em 1976, a técnica de sequenciamento químico. A metodologia de Sanger é fundamentada no princípio da terminação controlada da polimerização pela utilização de didesoxinucleotídeos, possibilitando assim determinar a ordem sequencial dos nucleotídeos que compõem fragmentos da molécula do DNA. Em sua versão original não era ferramenta apropriada para o sequenciamento de genomas complexos e no final da década de 80, após o método de Sanger ter sido modificado, o sequenciamento passou a ser realizado por meio automatizado e integrado a sistemas de leitura computadorizados (PEREIRA et al., 2013).

Atualmente plataformas baseadas em tecnologias de sequenciamento de nova geração (*Next Generation Sequence – NGS*), começaram a ser comercializadas e têm sido amplamente empregadas em substituição ao método de Sanger. Antes do sequenciamento propriamente dito, é necessário adotar estratégias para obtenção dos fragmentos de DNA do genoma a ser sequenciado, já que não é possível o sequenciamento de grandes segmentos de DNA. Para tal, são utilizadas duas estratégias principais: o sequenciamento *shotgun* de genoma inteiro (*whole genome shotgun – WGS*) e outra chamada de sequenciamento de clones ordenados. Ambas as estratégias visam gerar bibliotecas genômicas de clones, obtidas pela fragmentação do genoma em milhares de pequenos segmentos aleatórios, seguido da leitura de cada sequência e sobreposição destas, por meios computacionais. Com o auxílio de ferramentas de bioinformática, o NGS permite o sequenciamento e análise de milhões de pares de bases (pb) em rodada única de leitura (PEREIRA et al., 2013)

Introduzido em 2006, o sequenciador da Illumina é baseado no conceito de 'sequenciamento por síntese' (SBS) para produzir leituras sequenciais de aproximadamente 32-40 pb de dezenas de milhões de fragmentos de DNA amplificados à superfície simultaneamente. Partindo de uma mistura de fragmentos de DNA oligo-ligados adaptados de cadeia simples, o processo Illumina envolve o uso de uma estação de cluster microfluídica para adicionar esses fragmentos à superfície de uma célula de fluxo de vidro. Cada célula de fluxo é dividida em oito pistas separadas e as superfícies interiores têm oligos ligados covalentemente complementares aos adaptadores específicos que são ligados nos fragmentos da biblioteca (MARDIS, 2008).

A hibridização desses DNAs para os oligos na célula de fluxo ocorre por uma etapa ativa de aquecimento e resfriamento, seguida por uma incubação subsequente com reagentes e uma polimerase isotérmica que amplifica os fragmentos em uma área discreta ou "cluster" nas superfícies das células de fluxo. A célula de fluxo é colocada dentro do sequenciador, onde cada cluster é fornecido com polimerase e quatro nucleotídeos fluorescentes diferencialmente marcados que têm seu 3'-OH quimicamente inativado para garantir que apenas uma única base seja incorporada por ciclo. Cada ciclo de incorporação de bases é seguido por uma etapa de geração de imagens para identificar o nucleotídeo incorporado em cada cluster e por uma etapa química que remove o grupo fluorescente e desbloqueia a extremidade 3' para o próximo ciclo de incorporação de base. No final da execução do sequenciamento, a sequência de cada cluster é calculada e submetida a filtragem de qualidade para eliminar leituras de baixa qualidade (MARDIS, 2008).

2.2 Uso de ferramentas para montagem de genomas/metagenomas

Montagem é o processo de combinar sequências lidas em trechos contínuos de DNA, chamados *contigs*, baseados na similaridade das sequências lidas. A sequência consenso para um *contig* é o nucleotídeo de maior qualidade em qualquer leitura em cada posição, ou o nucleotídeo mais frequente em cada posição. O número de leituras em que a sequência consenso foi baseada é chamado de cobertura. Quando se está sequenciando um genoma completo, as

leituras são montadas em sequências ou *contigs* cada vez maiores, e finalmente no genoma completo. Na maioria dos metagenomas, uma montagem completa não é possível. Primeiro, porque a amostragem é incompleta e muitos, se não todos, os genomas das espécies estão apenas parcialmente amostrados. Segundo, porque a informação das espécies em si é incompleta, e é difícil mapear leituras individuais a suas espécies de origem. Existe também o perigo de se montar sequências a partir de leituras de espécies diferentes, gerando quimeras (WOOLEY; GODZIK; FRIEDBERG, 2010).

As quimeras são recombinantes artificiais entre duas ou mais sequências parentais. Essas moléculas artificiais dificultam a diferenciação da sequência original dos recombinantes, resultando em superestimação do nível de diversidade microbiana em amostras ambientais. Elas precisam ser identificadas e removidas do conjunto de dados usando ferramentas de bioinformática. Diversos algoritmos e ferramentas são utilizados para identificar preferencialmente sequências quiméricas em conjuntos de dados de alto rendimento (HUGENHOLTZ; HUBER, 2003).

A cobertura de um genoma é definida como o número médio de vezes que um nucleotídeo está sendo sequenciado. Comprimentos de leitura mais curtos (25 a 700, dependendo das tecnologias de sequenciamento) exigem mais cobertura, para garantir que todas as leituras se sobreponham e que essas sobreposições sejam únicas o suficiente para reconstruir o genoma através da montagem dos fragmentos. A cobertura geralmente é incompleta, já que a amostragem de sequência ambiental raramente produz todas as sequências necessárias para a montagem (WOOLEY; GODZIK; FRIEDBERG, 2010).

Na etapa de predição de genes a natureza incompleta e fragmentada dos dados metagenômicos apresenta desafios para a identificação de genes. Algoritmos de localização de genes são treinados para encontrar ORFs inteiras e levar em conta informações coletadas de grandes trechos genômicos. Para dados metagenômicos, no entanto, essas informações não estão disponíveis. Apesar dessas desvantagens, estudos mostraram que, para um conjunto de dados metagenômicos de alta complexidade, a predição de genes em montagens pode ser tão precisa quanto 85% dos genes originalmente previstos nos

genomas constituintes. Para um conjunto de baixa complexidade, isso vai até 90% (WOOLEY; GODZIK; FRIEDBERG, 2010).

A Ferramenta de Pesquisa de Alinhamento Local Básico (BLAST¹) localiza regiões de similaridade local entre sequências. O programa compara sequências de nucleotídeos ou proteínas com informações disponíveis em bancos de dados e calcula a significância estatística das correspondências. O BLAST pode ser usado para inferir relações funcionais e evolutivas entre sequências, bem como ajudar a identificar membros de famílias de genes. Esta abordagem informa a existência de membros da família de genes dentro de um metagenoma. O BLAST não pode ser usado para encontrar novas famílias e novos genes que não tenham homólogos em bancos de dados conhecidos. Para isso, são usadas ferramentas de predição de genes baseadas principalmente em métodos estatísticos de reconhecimento de padrões (BORATYN, 2012; KIM et al., 2013).

Após a predição de genes, é importante associá-los às suas espécies de origem (ou grupos taxonômicos). Essa análise é chamada *binning* ou classificação. Uma forma de classificar sequências é encontrar similaridades com sequências de referência que podem ser usadas para construir árvores. Essa técnica é útil quando a maioria das sequências da amostra possui similaridades significantes com sequências de referências conhecidas. Sequências de genes preditas, quando não tem homólogos, são adicionados em seu próprio nó isolado na árvore (WOOLEY; GODZIK; FRIEDBERG, 2010).

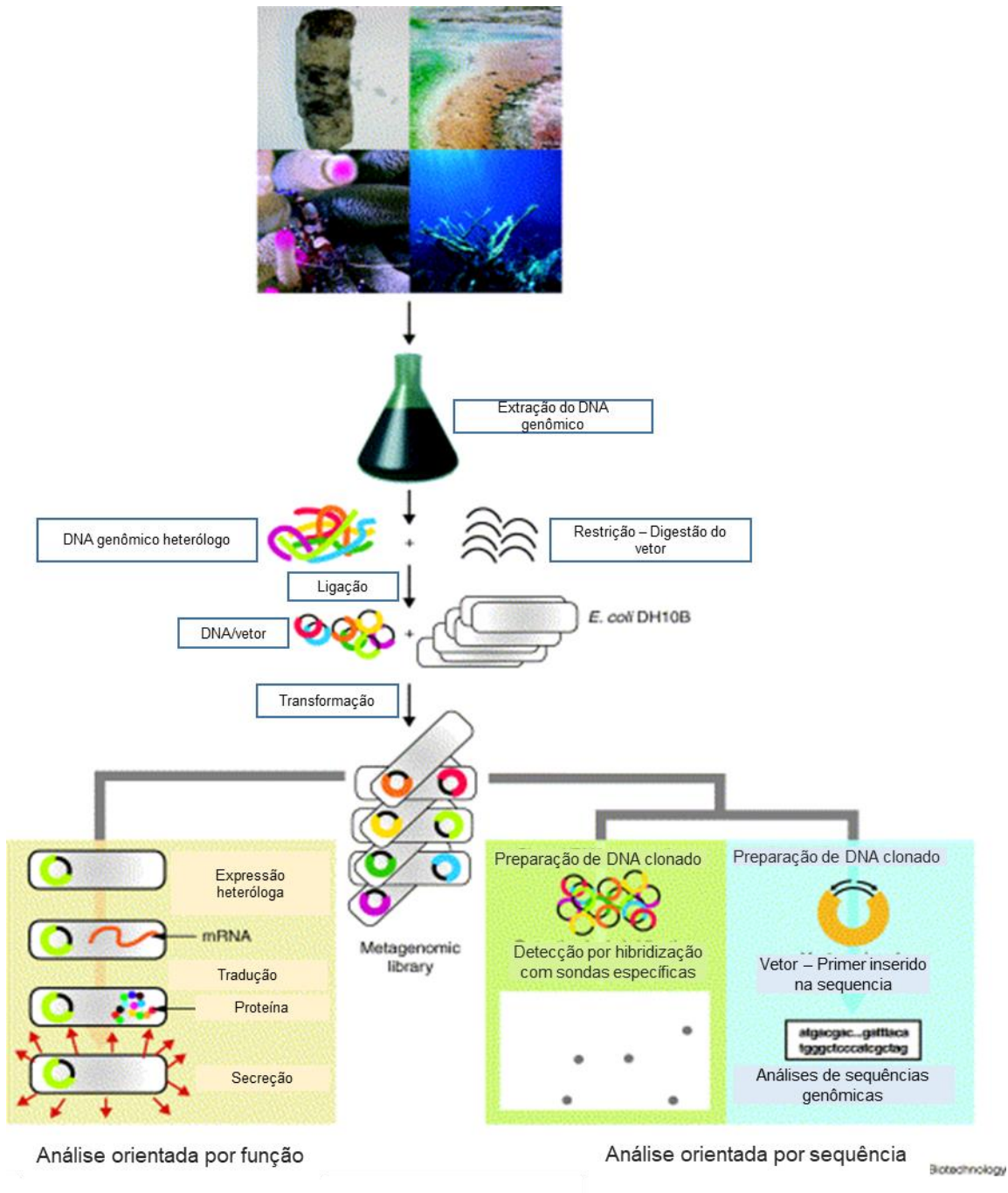
2.3 Estudo de amostras ambientais

Metagenômica é definida como o estudo de amostras ambientais, onde o DNA genômico total é analisado como um todo para identificação de microrganismos e genes em vias metabólicas, tendo como esse princípio uma das principais ferramentas de pesquisa em ecologia microbiana. A figura 1 ilustra de forma simplificada como é possível a construção de bibliotecas metagenômicas a partir de amostras ambientais. A informação metagenômica permite uma compreensão mais profunda do papel ecológico, do metabolismo e da história

¹ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, disponível em 4/6/2019.

evolutiva dos microrganismos em um dado ecossistema, analisando o DNA ambiental diretamente, sem cultivo prévio (KIM et al., 2013).

Figura 1. Construção e análise de bibliotecas metagenômicas a partir de amostras ambientais.



Fonte: Modificado de SCHLOSS e HANDELSMAN, 2003.

Embora a montagem de genomas microbianos completos tenha sido inicialmente restrita a amostras ambientais com baixa diversidade microbiana, aumentando a cobertura de sequenciamento, comprimentos de leitura e melhores algoritmos de agrupamento, se tornou possível a realização da montagem de genomas a partir de amostras de metagenomas, podendo estudar essas amostras através da montagem dos *contigs* usando algoritmos específicos, tornando possível o agrupamento de *contigs* e atribuição a grupos populacionais discretos (BOWERS et al., 2017).

É amplamente aceito que os procariotos possuem uma diversidade microbiana única e representam um reservatório biológico e genético amplamente inexplorado que pode ser estudado com o intuito de se descobrir novas enzimas com capacidades metabólicas únicas. No entanto, como já discutimos anteriormente, técnicas convencionais usadas em laboratório para cultura de bactérias frequentemente são ineficientes e limitados. Isso pode ser atribuído a diversos fatores, como alterações nos níveis de oxigênio atmosférico, condições osmóticas, nutrientes específicos necessários à sobrevivência, bem como condições de pH e temperatura. Assim, os ambientes de cultura de rotina fornecidos no laboratório impõem uma pressão seletiva que evitaria o crescimento da maioria dos microrganismos. Sendo as técnicas independentes de cultura essencial para entender a estrutura da população, os papéis ecológicos, a evolução e diversidade genética desses microrganismos encontrados em ambientes naturais (NGARA; ZHANG, 2018).

A metagenômica é uma derivação da genômica microbiana convencional, com a diferença significativa, que ela evita a necessidade de obter culturas puras para o sequenciamento. Portanto, a metagenômica mantém a promessa de revelar os genomas da maioria dos microrganismos que não podem ser facilmente obtidos em cultura pura. Sendo a técnica muito promissora para explorar os ricos recursos genéticos nos microrganismos não cultiváveis, ignorando o cultivo de microrganismos e isolando o DNA genômico diretamente de uma amostra ambiental (KUNIN et al., 2008).

A técnica metagenômica envolve duas abordagens básicas para triagem de biomoléculas a partir de amostras ambientais: triagem baseada em função e baseada em sequência de bibliotecas metagenômicas. Em ambos os casos, as

bibliotecas metagenômicas são construídas pela clonagem de DNA genômico fragmentado em vetores de expressão apropriados, tais como plasmídeo, cosmídeo, fago lambda ou foscídeos que permitem a expressão dos genes desejados (COUGHLAN et al., 2015).

O rastreo baseado em sequência é utilizado com maior frequência do que as abordagens baseadas em funções, devido ao fácil acesso aos dados de sequenciamento metagenômico e a muitos *softwares* disponíveis para análise de dados. No entanto, abordagens baseadas em sequências têm desvantagens inerentes, com sua eficácia amplamente afetada pela anotação do genoma e completude dos dados disponíveis. Essas abordagens respondem aos algoritmos disponíveis e às informações presentes nos bancos de dados para inferir as funções dos genes recém-descobertos. Assim, eles podem não funcionar bem se a similaridade da sequência não corresponder a uma relação funcional, ou se o novo gene tiver apenas uma semelhança fraca com quaisquer genes cujos produtos tenham sido examinados bioquimicamente, ou se um gene específico for capaz de realizar numerosas funções na célula (NGARA; ZHANG, 2018).

Bibliotecas de genes metagenômicos (Figura 1) podem ser usadas para a identificação de novas enzimas codificadas por um único gene ou operon de tamanho pequeno, enquanto bibliotecas de tamanho grande são necessárias para o isolamento de grandes aglomerados de genes biossintéticos, que codificam caminhos complexos contendo vários genes. Após a construção da biblioteca, os clones metagenômicos resultantes são usados para transformar um hospedeiro heterólogo, que na maioria dos casos é a bactéria *Escherichia coli* (COUGHLAN et al., 2015).

As análises metagenômicas começam com o isolamento do DNA microbiano de uma amostra ambiental. O espécime de DNA metagenômico adquirido deve ser tão puro e de alta qualidade quanto possível, e deve representar com precisão todas as espécies presentes qualitativa e quantitativamente. O sequenciamento direto do DNA metagenômico extraído, seguido de análises bioinformáticas adequadas, ajudam na elucidação das características funcionais de microrganismos colonizadores de ambientes particulares. Através do sequenciamento do gene 16S rRNA, um perfil funcional das bactérias presentes em um dado ambiente pode ser obtido. Informações sobre os papéis funcionais de espécies bacterianas já

estudadas estão disponíveis em arquivos de bancos de dados, incluindo tanto bactérias cultivadas cujas proteínas funcionais foram extensivamente caracterizadas quanto funções atribuídas a proteínas bacterianas produzidas por bactérias não cultivadas através de estudos prévios. Uma vez que um membro de uma família bacteriana previamente descrita tenha sido identificado em uma amostra ambiental, ou um parente conhecido mais próximo tenha sido designado, a análise filogenética pode atribuir funções previstas a uma espécie bacteriana identificada, referindo-se à informação funcional disponível em relação a esse grupo taxonômico específico (BOWERS, 2017).

O crescimento do uso da tecnologia *Next Generation Sequencing* (NGS), pode ser visto pela quantidade de plataformas disponíveis para sequenciamento no mercado, dentre as mais utilizadas para estudos metagenômicos destacam-se o Miseq/HiSeq (Illumina), PacBio, Nanopore, onde a combinação da metagenômica funcional e ferramentas computacionais vem auxiliando em diversos estudos de resistomas (CROFTS; GASPARRINI; DANTAS, 2017; SURETTE; WRIGHT, 2017).

A mineração de dados genômicos ou metagenômicos vem sendo tendência para permitir e acessar a informação para selecionar novos genes de interesse em biotecnologia, utilizando mecanismos de busca de genes como por exemplo o *screening* virtual, técnica utilizada para buscar informações em dados de sequenciamento, que vem auxiliando pesquisadores a identificarem novas proteínas, mecanismos, estudos metabólicos, compostos secundários gerados, alimentando dessa maneira novos métodos computacionais para a estruturação e previsão de sua função (SHARPTON, 2014; CROFTS; GASPARRINI; DANTAS, 2017).

3. OBJETIVOS

Realizar o estudo metagenômico de uma comunidade bacteriana para prospecção de sequência codificadora de arabinofuranosidase, e avaliar o potencial biotecnológico da enzima a ser expressa, sendo essa pertencente ao genoma de bactéria do gênero *Chitinophaga*, isolada inicialmente em uma pilha de bagaço de cana-de-açúcar.

4. MATERIAL E MÉTODOS

Sequências codificadoras da enzima α -L-arabinofuranosidase foram selecionadas do banco de dados (CB10) (<http://200.145.102.111/cb10/>) construído a partir do sequenciamento de um consórcio microbiano. Essas sequências foram prospectadas por similaridade e analisadas *in silico*, sendo elas CB10_2.116, CB10_61.2456 e CB10_282.6219. Desse estudo foi possível selecionar a sequência que apresentou melhores resultados, sendo a sequência nomeada CB10_2.116, a escolhida para dar continuidade aos testes de expressão e caracterização bioquímica conforme resultados apresentados no próximo capítulo. Neste capítulo faremos uma breve descrição das características da sequência escolhida. As sequências foram cedidas pelo Dr. Luciano Takeshi Kishi do Laboratório de Bioquímica de Microrganismos e Plantas (LBMP) da UNESP/FCAV de Jaboticabal, que possui o genoma de uma *Chitinophaga* sp (KISHI et al., 2017).

A comunidade bacteriana foi isolada de uma pilha de bagaço de cana-de-açúcar e pertence à coleção de microrganismos do Laboratório de Bioquímica de Microrganismos e Plantas da faculdade de Ciências Agrárias e Veterinárias da UNESP (LBMP).

Uma alíquota de 100 μ L do estoque de uma comunidade bacteriana (CB10) em glicerol 20% foi adicionada a 900 μ L de solução salina 0,85% (m/v) e submetida a diluições seriadas até 10^{-9} . Foram espalhadas 100 μ L de cada diluição em placas de Petri contendo meio de cultura sólido PGE (0,5 g.L⁻¹ K₂HPO₄; 0,2 g.L⁻¹ MgSO₄ 7.H₂O; 0,1 g.L⁻¹ NaCl, 1,0 g.L⁻¹ de extrato de levedura, 10,0 g.L⁻¹ de glicose e 9,0

g.L⁻¹ de ágar, pH 6,8). As placas foram incubadas a 30°C durante 72 horas. Após o período de incubação as colônias de bactérias foram isoladas com base nas características morfológicas relacionadas ao seu tamanho, forma, elevação, cor e textura. Os isolados selecionados foram cultivados isoladamente no mesmo meio de cultura sob as mesmas condições e usados nos ensaios microbiológicos, bioquímicos e moleculares.

4.1 Extração de DNA total e sequenciamento da comunidade bacteriana CB10.

A comunidade CB10 foi cultivada em meio PGE líquido durante 36 horas, sob agitação constante a 150 rpm a 30°C. A partir da suspensão bacteriana obtida, foi transferida uma alíquota de 1000 µL para um microtubo de 1,5 mL que foi processado em centrífuga Eppendorf® modelo 5810R a 12.000 xg por 20 minutos. Após a centrifugação o sobrenadante foi descartado e o precipitado foi usado para a extração de DNA. O DNA total da comunidade (CB10) foi extraído utilizando kit da PROMEGA – Wizard® *Genomic DNA Purification* de acordo com as recomendações do fabricante.

A quantificação do DNA total da comunidade CB10 foi realizada por espectrofotometria no comprimento de onda 260/280 nm (NanoDrop® ND-1000 UV-Vis) e por Fluorometria Qubit® dsDNA BR (Life Technologies Europa, Naerum, Dinamarca). A qualidade do DNA foi avaliada em eletroforese em gel de agarose 0,8% (m/v) contendo brometo de etídeo (10 µg.mL⁻¹). A biblioteca do DNA metagenômico (CB10b) foi preparada usando o kit da Nextera® XT DNA Library e posteriormente foi submetido ao sequenciamento em uma “lane” em HiSeq Flow Cell v4, com o HiSeq SBS Kit v4 e leituras pareadas de 2 x 100 bp na plataforma Illumina HiSeq 2500.

4.2 Montagem, classificação taxonômica e perfil funcional da comunidade bacteriana (CB10).

Os dados brutos gerados pelo sequenciamento foram processados através do programa Atlas v1.0.2 (WHITE III et al., 2017), analisado quanto à qualidade,

sendo retirados eventuais segmentos de sequências contaminantes de regiões de baixa qualidade usando BBDuk que pertence ao pacote de ferramentas do BBMap (<https://sourceforge.net/projects/bbmap/>). Os dados que passaram pela etapa de qualidade foram normalizados quanto à cobertura de leituras com base na frequência de k-mer formando subconjuntos de leituras usando a ferramenta BBNorm. Os subconjuntos com sequências de alta qualidade foram usados como entrada para o montador SPAdes v3.10.1 (BANKEVICH et al., 2012).

Após a montagem, os contigs foram submetidos para a predição das ORFs (*Open Reading Frame*) pelo programa Prodigal (HYATT et al., 2010). As proteínas preditas foram anotadas usando alinhamento de proteína-proteína através do DIAMOND (BUCHFINK; XIE; HUSON, 2015), com filtro padrão > 200 para o bitscore e e-value de < $1e^{-7}$ para os bancos do RefSeq (LEARY et al., 2016), EggNOG (HUERTA-CEPAS et al., 2016), dbCAN para famílias CAZy (YIN et al., 2012), ENZYME para comissão enzimática (EC) (BAIROCH, 2000) e COG (TATUSOV et al., 2003), implementados na plataforma do programa Atlas (WHITE III et al., 2017).

4.3 Avaliação do potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica

A anotação funcional de ORFs com domínios associados às enzimas ativas em carboidratos (CAZy) foi realizada por busca de similaridade de domínios relacionados às famílias do CAZy com base em modelos ocultos de Markov (HMMs) no dbCAN (<http://csbl.bmb.uga.edu/dbCAN>) (YIN et al., 2012). Foram aplicados os parâmetros de 30% de cobertura do HMM, e-value de $1e^{-5}$ para alinhamentos maiores que 80 aminoácidos e de $1e^{-3}$ para alinhamentos menores que 80 aminoácidos. As enzimas preditas com domínios para carboidratos ativos no metagenoma foram adicionalmente comparadas contra os bancos de dados não redundante do GenBank (NCBI), UNIProt (APWEILER, 2009) e o InterProscan (ZDOBNOV; APWEILER, 2001) considerando os parâmetros e-value < 10^{-5} e um valor mínimo de 50% de identidade para inferir função às proteínas putativas.

As análises de montagem e anotação de genoma foram previamente realizadas Dr. Luciano Takeshi Kishi Faculdade de Ciências Agrárias e Veterinárias

– Unesp, Campus de Jaboticabal o que deu origem as sequencias que se trata este trabalho (KISHI et al., 2017).

4.4 Alinhamento de múltiplas sequências

Para o alinhamento de múltiplas sequências (MSA), importante nas análises evolutivas de sequências biológicas utilizou-se o MAFFT que é um programa de MSA, lançado pela primeira vez em 2002 (KATO et al. 2002). O MAFFT versão 7 tem opções para várias estratégias de alinhamento, incluindo métodos progressivos, métodos de refinamento iterativo e métodos de alinhamento estrutural para RNAs. Existem duas classes significativamente diferentes de métodos MSA, métodos baseados em similaridade e métodos baseados em evolução. O MAFFT é classificado como um método baseado em similaridade. No entanto, a informação evolutiva é útil mesmo para métodos baseados em similaridade, porque as sequências a serem alinhadas são geradas a partir de um ancestral comum no curso da evolução. A este respeito, o MAFFT leva em conta informações evolutivas (KATO; STANDLEY, 2013).

4.5 Software de Análise Genética Evolutiva Molecular

Os numerosos passos necessários para colher resultados interpretáveis e acionáveis a partir de dados de sequência bruta invariavelmente requerem uma análise comparativa de sequências moleculares para descobrir diferenças funcionais e adaptativas do genoma. O software de Análise Genética Evolutiva Molecular (MEGA) fornece ferramentas para conduzir tais análises. O MEGA (<https://www.megasoftware.net/>) inclui um vasto repertório de programas para a montagem de alinhamentos de sequências, inferindo árvores evolutivas, estimando distâncias genéticas e diversidades, inferindo sequências ancestrais, horários de cálculo e seleção de testes (KUMAR et al., 2012).

4.6 Construção de Árvores filogenéticas

O DoMosaics é um programa JAVA e requer linguagem JAVA 1.6 ou mais atual. Árvores filogenéticas podem ser construídas com base em sequências remotas usando serviços SOAP ou localmente usando distâncias entre arranjos de domínio na forma de edição de domínio ou distâncias de Jaccard. As pesquisas por arranjos de domínio semelhantes por alinhamentos de cadeia de domínio são realizadas usando o serviço da Web RADS/RAMPAGE. Finalmente, DoMosaics permite a anotação dependente de conteúdo. O programa pode ser iniciado diretamente da página da web DoMosaics usando o JAVA Web Start. Alternativamente, um arquivo JAVA pré-compilado pode ser baixado (MOORE et al., 2014).

Já o serviço da web SyntTax utilizado, tem como objetivo auxiliar a análise da sintenia ("mesma fita" em grego, tal termo foi introduzido para definir loci posicionados no mesmo cromossomo, independentemente de estarem geneticamente ligados ou não), fornecendo uma árvore organizada que expõe a linhagem completa de todos os procariontes completamente sequenciados no repositório do Centro Nacional de Informações sobre Biotecnologia (NCBI). Desta forma, organismos simples ou múltiplos podem ser facilmente selecionados por qualquer combinação de classificações, permitindo a geração de sintenias robustas, realizadas de acordo com critérios taxonômicos. Os bancos de dados taxonômicos e genômicos são armazenados localmente no servidor SyntTax; ambos são atualizados automaticamente diariamente a partir dos respectivos recursos do NCBI. Esse recurso faz do SyntTax uma ferramenta flexível, capaz de se adaptar prontamente à adição de novos genomas sequenciados (OBERTO, 2013).

5. RESULTADOS E DISCUSSÃO

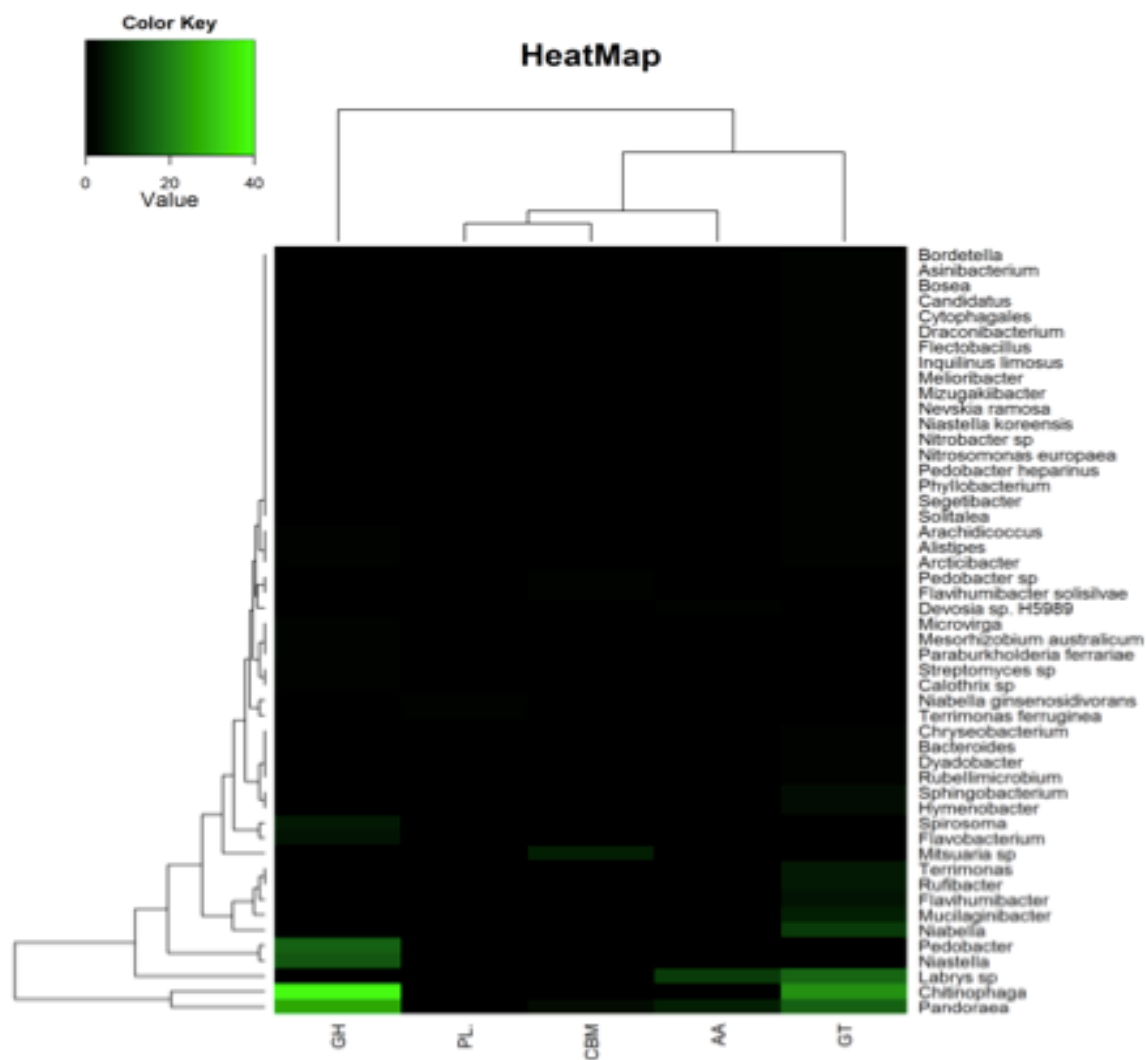
A metagenômica é uma área relativamente jovem que tem ampla aplicação em biologia e biotecnologia. Embora muitos avanços na expressão gênica heteróloga, na construção de bibliotecas, no desenho vetorial e na triagem possam

melhorá-la, a tecnologia atual é suficientemente poderosa para gerar produtos para resolver problemas do mundo real, incluindo a descoberta de novos produtos biotecnológicos. Abordagens que enriquecem uma parte da comunidade microbiana ou uma coleção de clones metagenômicos aumentarão o poder da análise metagenômica para abordar questões específicas da ecologia microbiana e descobrir novas aplicações biotecnológicas (SCHLOSS; HANDELSMAN, 2003).

Muitas enzimas foram descobertas por meio da metagenômica, incluindo lignases, xilanases, endoglucanases, enzimas amilolíticas, β -glicosidasas para bioetanol e enzimas lipolíticas para biodiesel. Algumas delas têm múltiplas funções e podem catalisar várias reações diferentes, podem exibir atividades elevadas, especificidades, estabilidade e ainda podem trabalhar sob uma ampla gama de pH, temperatura ou força iônica. Todas essas características fazem das enzimas importante fonte de estudo, se tratando de uma alternativa para aplicação na produção de biocombustíveis. As enzimas que participam da degradação de biomassa identificadas a partir de amostras ambientais incluem: carboxil-hidrolases (esterases, lipases), enzimas modificadoras de polissacarídeos (celulases, α -amilases, xilanases, enzimas ramificadoras de 1, 4- α -glucanas), oxidorreduções, desidrogenases e oxigenases (XING; ZHANG; HUANG, 2012).

No mapa térmico ilustrado na figura 2 podemos visualizar a abundância de famílias de enzimas como Glicosil hidrolases, Glicosil transferase, Polissacarídeo liases, CBM e enzimas auxiliares, sendo que as espécies descritas fazem parte do consórcio microbiano isolado a partir de amostras de bagaço de cana dando origem ao genoma CB10, vale ressaltar que o gênero *Chitinophaga* foi o representado com maior quantidade de cazy-enzimas especialmente de GHs.

Figura 2. Mapa térmico mostrando abundância para GH, GT, PL, CBM e AA no consórcio que deu origem ao genoma CB10, sendo que *Chitinophaga* foi o gênero representado com maior quantidade de cazy-enzimas. Famílias com alta abundância são mostrados na cor verde. O agrupamento hierárquico no eixo horizontal é baseado na abundância média de cada família CAZy incluída na análise. O agrupamento hierárquico do eixo vertical baseia-se na similaridade entre a distribuição dessas CAZy entre os grupos de organismos.



Fonte: autoria própria.

A classificação das enzimas em famílias é feita de acordo com sua função no processo de clivagem ou construção de hidratos de carbono complexos, sendo as glicosil hidrolases (EC 3.2.1.-) um grupo disseminado de enzimas que hidrolisam a ligação glicosídica entre dois ou mais carboidratos ou entre uma porção de carboidrato e não-carboidrato (FLORENCIO; BADINO; FARINAS, 2017).

Lembrando que os avanços sustentáveis na degradação da parede celular das plantas dependem profundamente de enzimas ativas em carboidratos. Várias dessas atividades enzimáticas atrasam-se no que diz respeito à compreensão das relações estrutura/função, bem como fontes biológicas de enzimas com potencial de aplicação proeminente (WILKENS et al., 2017). Com isso torna-se evidente a importância de se estudar as fontes ambientais que contêm tais enzimas e acima de tudo buscar desvendar as suas aplicações e funções.

A lignocelulose representa mais de 90% do peso seco de uma célula vegetal, sendo composta pelos polímeros celulose, hemicelulose e lignina, unidos fortemente entre si por forças não covalentes e ligações covalentes nas paredes celulares de tecidos vasculares de plantas terrestres superiores. A quantidade de cada um dos polímeros pode variar de acordo com cada espécie e a idade da planta, bem como entre as partes de uma planta. Em média, a lignocelulose consiste de 45% de celulose, 30% de hemicelulose e 25% de lignina. As lignoceluloses são os compostos orgânicos mais abundantes na biosfera abrangendo aproximadamente 50% da biomassa no mundo, com uma produção anual estimada entre 10 e 50 x 10⁹ toneladas. Esta produção é referente as atividades da agricultura, florestas, frutas e vegetais. (MENEZES; BARRETO, 2015).

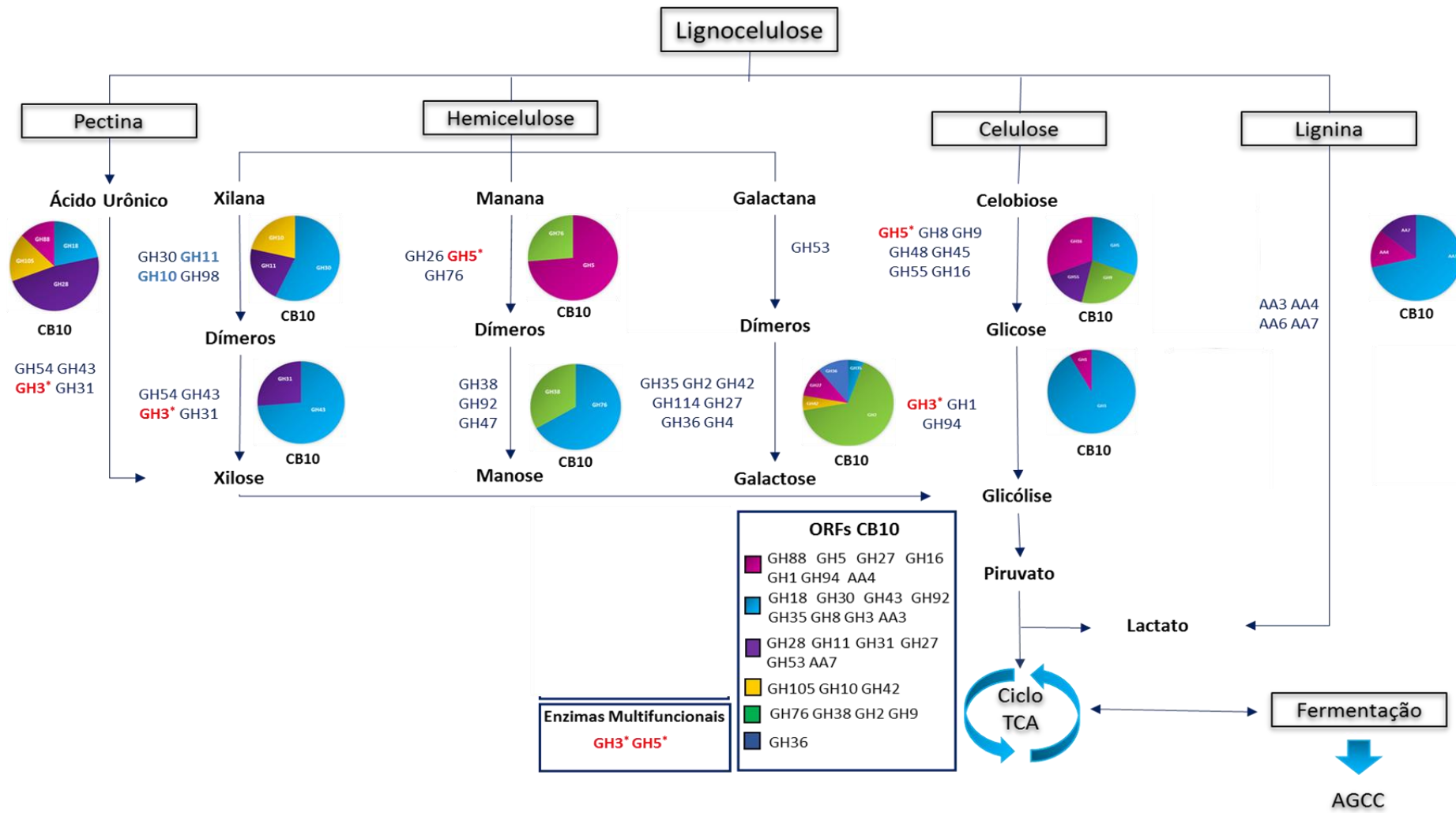
As enzimas que degradam a xilana possuem um grande potencial em várias aplicações na área da biotecnologia, podemos observar que na figura 3 que pelo menos 8 famílias de enzimas podem atuar na degradação da cadeia de xilose. As xilanases estão envolvidas na bioconversão da xilana em xilose, o qual age como substrato na obtenção de etanol através de processos de fermentação e em xilitol, utilizado como adoçante em vários tipos de alimentos. O interesse nas xilanases aumentou, devido ao seu uso em processos de branqueamento da polpa de papel, com processos que dispensam a utilização do cloro (MENEZES; BARRETO, 2015).

Outra família de enzimas que tem atraído grande interesse são as arabinofuranosidases pois são importantes na degradação eficiente e completa das hemiceluloses e pectinas da parede celular vegetal e a sua aplicação tem benefícios econômicos e ambientais. Wilkens et al. (2017) relata em seu trabalho que antes da fermentação, a hemicelulose é pré-tratada por hidrólise ácida ou submetida a tratamento alcalino gerando subprodutos de oligossacarídeos não

fermentáveis que inibem a fermentação. Uma solução alternativa promissora consiste na degradação enzimática das hemiceluloses utilizando um coquetel de enzimas arabinofuranosidases, hemicelulolíticas, xilanases, esterases e xilosidases, que promove uma significativa conversão da hemicelulose em monossacarídeos.

Uma diversidade de genes envolvidos na degradação da lignocelulose são relevados quando observamos a figura 3, esses foram identificados como Cazyne para CB10 –*Chitinophaga* sp. Dentre as famílias de GHs podemos destacar algumas envolvidas na degradação de α -L-arabinofurano-oligossacarídeos, entre essas podemos citar famílias GH3, GH43 e GH54. Além disso, as hidrolases glicosídicas nas famílias GH42 e GH27 possuem atividades de α e β -L-arabinopiranosidase. Em um estudo publicado por Fujita et al. (2011) descreveram a nova β -L-arabinobiosidase, tratando-se do primeiro relato de uma enzima que hidrolisa os β -L-arabinofuranosídeos.

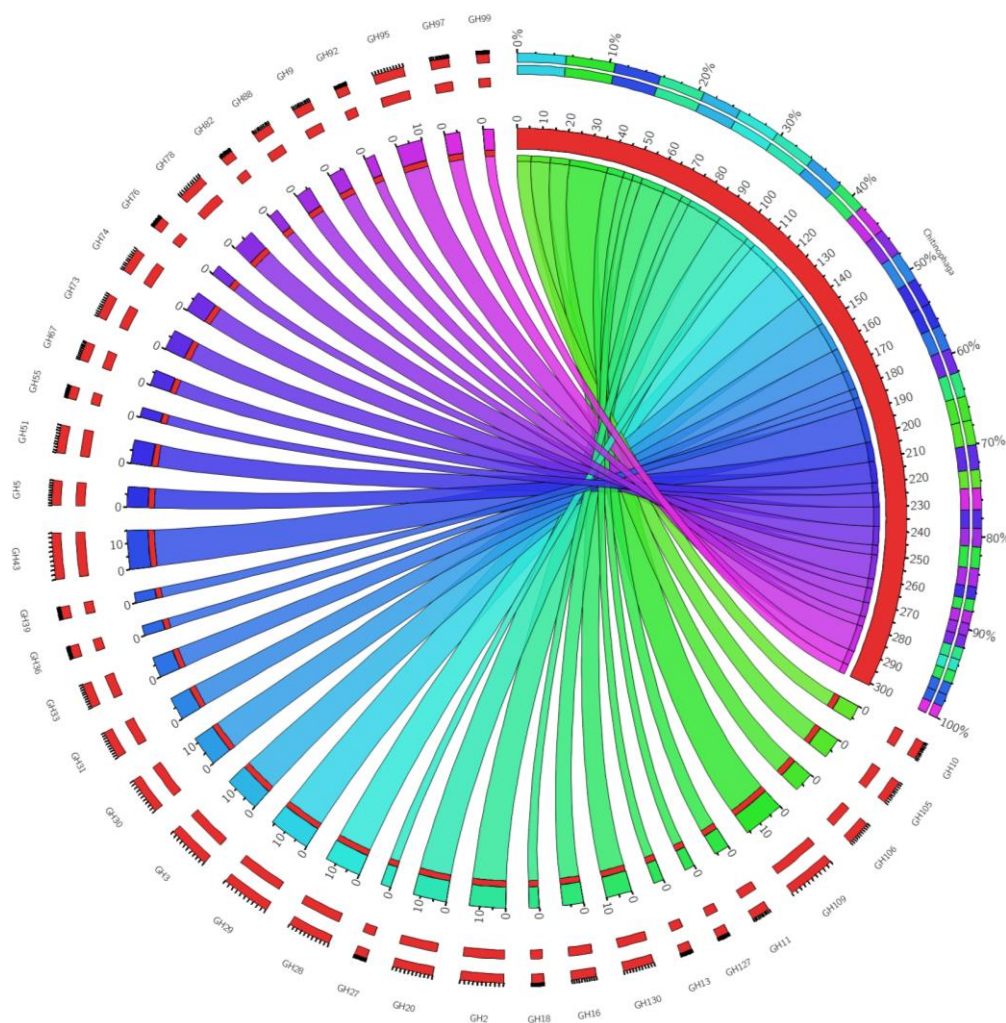
Figura 3. Representação dos genes envolvidos em degradação de biomassa identificados como Cazy-enzimas para CB10 –*Chitinophaga* sp.. Comparação da abundância de CAZyme expressas encontradas na comunidade com abundância de microrganismos identificados.



Fonte: Autoria própria

Na análise realizada nesse estudo é representada pela figura 4, podemos observar a distribuição de ORFs relacionadas as famílias de GHs. As ORFs preditas para o genoma de *Chitinophaga* sp. foram comparadas com enzimas do banco de dados do CAZy, apresentando um total de 37 ORFs associadas a enzimas ativas para carboidratos. Dessas as famílias mais representativas são as GH2, GH28, Gh29, GH43 e GH109. Diante disso se pode observar a infinidade de genes responsáveis por expressar enzimas relevantes do ponto de vista biotecnológico, sendo esses revelados a partir de estudo metagenômicos. Essa análise teve como finalidade apresentar o potencial metabólico da comunidade bacteriana para desconstrução da biomassa lignocelulósica.

Figura 4. Circos da distribuição de ORFs para GHs no genoma de *Chitinophaga* sp. A largura das barras de cada família enzimática indica sua abundância relativa no genoma.



Fonte: Autoria própria.

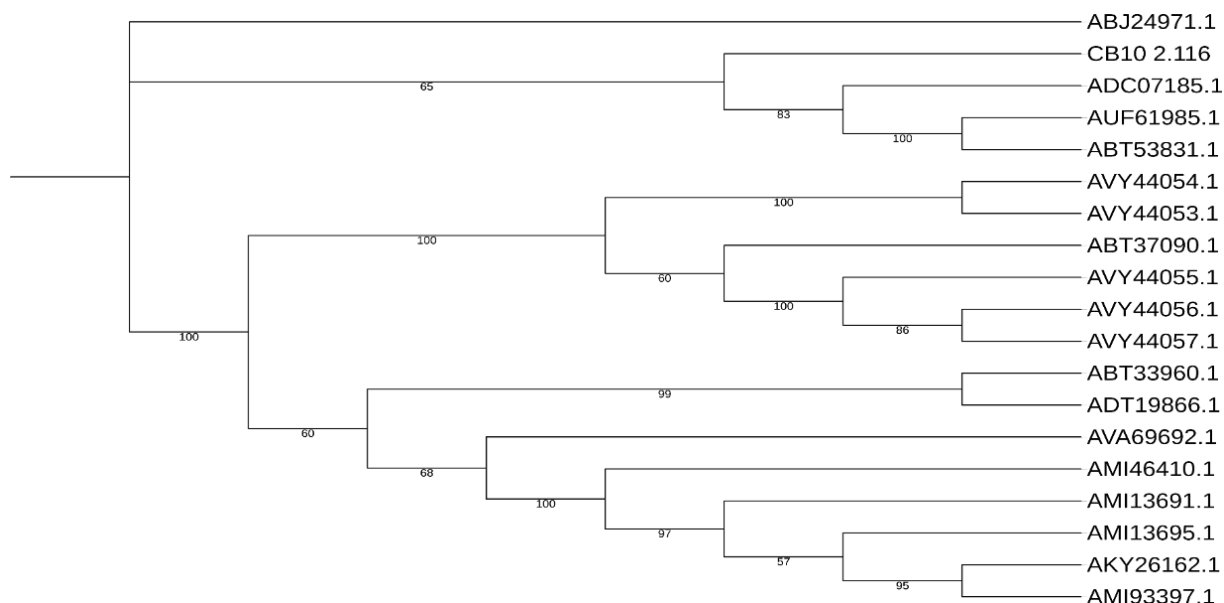
5.1 Caracterização da ORF selecionada CB10 2.116 que codifica uma arabinofuranosidase.

No âmbito da biotecnologia, um sistema de patentes, articulado a outras ações de incentivo, e uma cultura de propriedade intelectual, estimulam o crescimento econômico. Em 2010, os negócios relativos à biotecnologia movimentaram 84,6 bilhões de dólares o que, sem dúvida, pressiona países a adotarem políticas liberais quanto a pesquisas e experimentações biotecnológicas (NAVES, 2013).

Sendo assim se faz necessário verificar se o material de estudo já possui algum registro em bancos de patentes. Conforme podemos observar na figura 5, a análise fenética aponta que a sequência CB10 2.116 referente a família GH127 quando comparada na base de dados de patentes não possui até o momento registro de patente, uma vez que a mesma está localizada em um ramo mais distante das sequências já patenteadas. O gene foco do estudo é mais aparentado a sequências de um ramo e mais distante do grupo maior. Dessa forma se torna possível o estudo de tal sequência com a intenção de investigar sua funcionalidade e propriedades que sejam interessantes a aplicações biotecnológicas.

Outra análise foi realizada a fim de comparar a sequência de estudo com outras sequências codificadoras de enzimas, dessa vez, a comparação foi feita analisando sequências já depositadas na base de dados CAZy. O banco de dados CAZy descreve as famílias de módulos de ligação catalítica e de carboidrato estruturalmente relacionados (ou domínios funcionais) de enzimas que degradam, modificam ou criam ligações glicosídicas.

Figura 5. Dendrograma hierárquico obtido para sequências de patentes de GH127 depositadas na base de dados de patentes. O dendrograma foi obtido pelo método de otimização para vizinhos próximos (*neighbor joining*).



Fonte: Autoria própria.

Online desde 1998, o CAZy é um banco de dados especializado dedicado à exibição e análise de informações genômicas, estruturais e bioquímicas sobre enzimas ativas de carboidratos. Novos genomas são adicionados regularmente logo após aparecerem nos lançamentos diários do GenBank. Novas famílias são criadas com base em evidências publicadas para a atividade de pelo menos um membro da família e todas são atualizadas regularmente, tanto em conteúdo quanto em descrição. (<http://www.cazy.org/>).

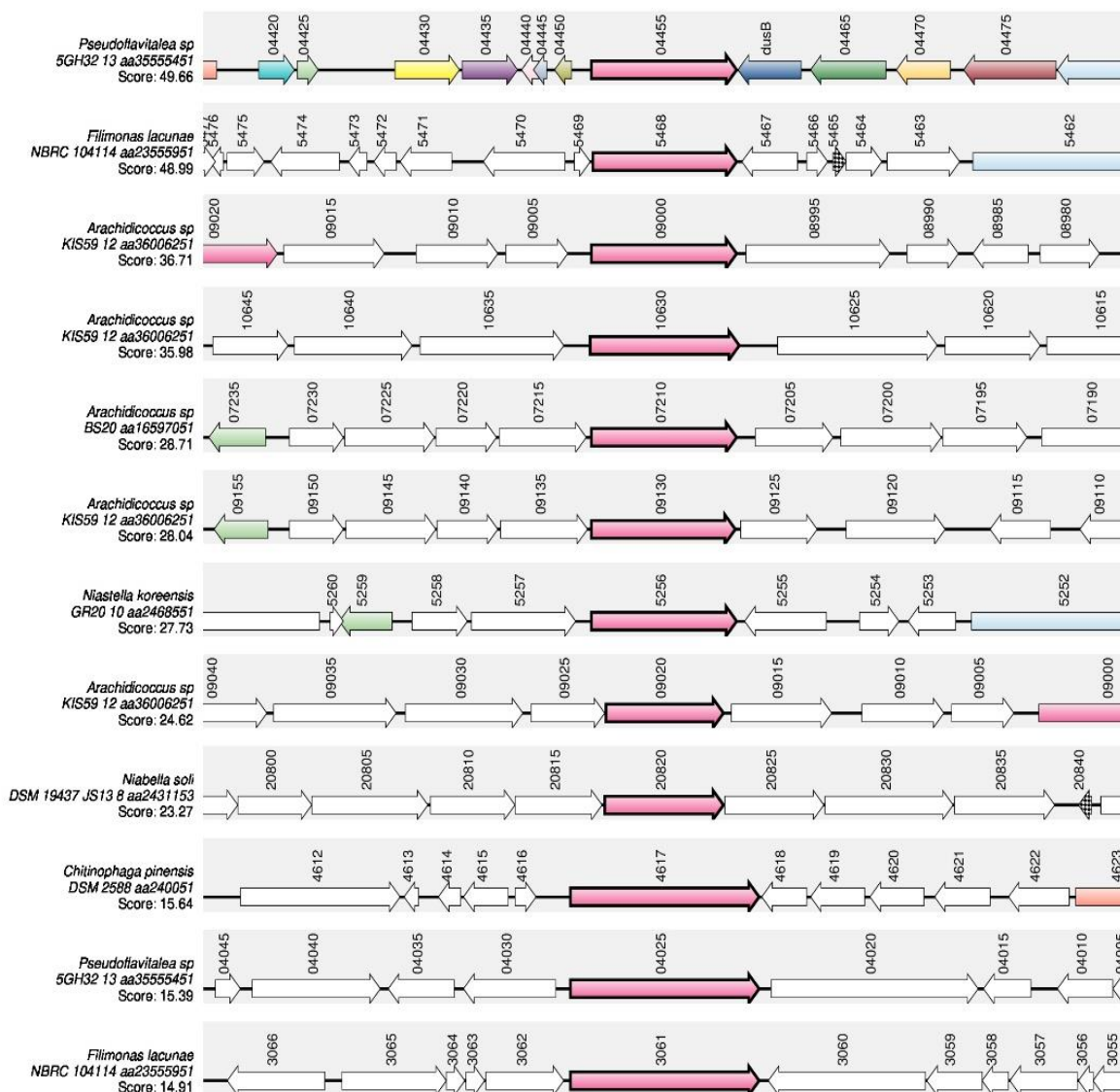
Mais uma vez podemos notar que a sequência CB10 2.116, está localizada em um ramo mais distante e isolada das outras sequências já caracterizadas, comprovando dessa maneira ser uma sequência inédita validando os esforços na caracterização da mesma. Os resultados podem ser visualizados na figura 6.

5.2 Expressão de genes responsáveis pela produção de enzimas degradadoras de lignocelulose.

Muitos microrganismos expressam genes responsáveis pela produção de enzimas capazes de fazer a bioconversão de substratos lignocelulósicos em compostos de fácil assimilação para o seu metabolismo, e sabemos que as enzimas hidrolíticas tem papel fundamental nessa bioconversão agindo sinergicamente formando um complexo com várias enzimas (MENEZES; BARRETO, 2015).

Estas enzimas principalmente as que hidrolisam ligação glicosídica são agrupadas em famílias de glicosil hidrolases (GH) com base nas semelhanças de sequência de aminoácidos (<http://www.cazy.org>). A figura 7 ilustra um comparativo entre os diferentes microrganismos e a presença de genes com a mesma funcionalidade que a sequência CB10 2.116, como exemplo podemos destacar presença do gene na bactéria *Chitinophaga pinensis*, um parente próximo do gênero foco de estudo nesse trabalho. Um estudo apresentado por Rio et al. (2010), para o genoma de *C. pinensis* são previstas glicosil hidrolases pertencentes a 49 famílias de GH diferentes; 18 delas pertencentes à família GH 43 que contém xilosidases, xilanases, arabinanases, arabinofuranosidases e galactosidases.

Figura 7. Análise sintênica do gene β -L-arabinofuranosidase, ilustração da comparação entre os genes de diferentes microrganismos que expressão tal enzima. (Comparações foram obtidas com a ferramenta online SyntTax - <http://archea.u-psud.fr/syntax/Default.aspx>).

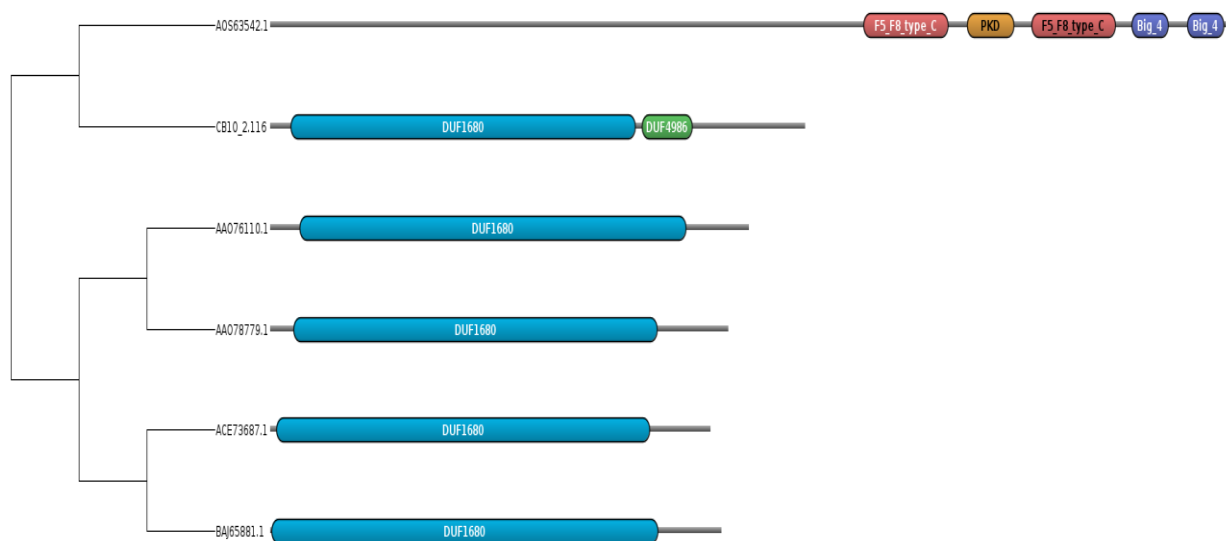


Fonte: Autoria própria.

A figura 8 apresenta um esquema comparativo entre GHs já caracterizadas evidenciando a disposição dos domínios proteicos. Conforme se pode observar a sequência CB10_2.116 se apresenta mais aparentada a uma sequência pertencente à família GH 121.

A sequência AOS63542.1 é descrita no CAZY como pertencente a ordem *Actinomycetales* que conforme descrito por Schaffert et al. (2016) essas são consideradas uma fonte importante de uma variedade de novos compostos bioativos e possivelmente farmacologicamente importantes e drogas, tais como agentes anticâncer, antibióticos e também outras moléculas industrialmente relevantes e enzimas com diversas atividades biológicas. Especialmente os actinomicetos marinhos se tornaram um foco de pesquisa, uma vez que desenvolveram a maior diversidade genômica e metabólica e são fontes promissoras de novos metabólitos e enzimas.

Figura 8. Dendrograma hierárquico obtido para sequências de GHs 121, 127 e 125 caracterizadas e depositadas na base de dados Cazy dendrograma foi obtido pelo método de “otimização” para vizinhos próximos (*neighbor joining*) juntamente com mapeamento de domínios junto ao programa DoMosaics.



Fonte: Autoria própria.

Diante das análises apresentadas nesse capítulo foi possível observar que a sequência CB10_2.116 pode ser uma interessante candidata a estudos de caracterização, uma vez que quando comparada a outras sequências já

caracterizadas e funcionais, do ponto de vista das análises fenética, se mostrou aparentada a outras enzimas com interesse biotecnológico, sendo essa uma importante candidata a estudos com foco em enzimas de interesse industrial. O próximo capítulo apresenta uma descrição de três sequencias, obtidas atrás do sequenciamento do genoma de *Chitinophaga*, juntamente com a clonagem e expressão da sequência CB10_2.116.

6. CONCLUSÕES PARCIAIS

Conforme exposto neste capítulo foi possível notar que nos últimos anos, a metagenômica acelerou o entendimento da microbiologia, graças aos avanços técnicos das plataformas de sequenciamento. Aqui utilizamos dessas ferramentas para compreender a sequência em estudo e através de análises comparativas com outras sequências já caracterizadas comprovar que a sequência CB10_2.116 pode ser uma potencial enzima para estudos voltados a área biotecnológica, uma vez que atende a requisitos como, não estar depositada em bancos de patentes, ser aparentada a enzimas já caracterizadas e que apresentam interesse biotecnológico.

REFERÊNCIAS

- ALMEIDA, M. L. M. Diversidade bacteriana de um consórcio degradador de resíduo celulósico. Faculdade de Tecnologia de Jaboticabal, 2011.
- APWEILER, R. The universal protein resource (UniProt) in 2010. **Nucleic Acids Research**, v. 38, n. SUPPL.1, p. 190–195, 2009.
- BAIROCH, A. The ENZYME database in 2000. **Nucleic acids research**, v. 28, n. 1, p. 304–5, 2000.
- BANKEVICH, A. et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012.
- BARBOSA, E. B. et al. Proteômica: metodologias e aplicações no estudo de doenças humanas. **Rev Assoc Med Bras**, v. 58, n. 3, p.366-375, 2012.
- BORATYN, G. M et al. Domain enhanced lookup time accelerated BLAST. **Biology Direct**, v. 7, n. 1, p.7-12, 2012.
- BOWERS, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. **Nature Biotechnology**, v. 35, n. 8, p.725-731, 2017.
- BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature methods**, v. 12, n. 1, p. 59–60, 2015.
- COUGHLAN, L. M. et al. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. **Frontiers In Microbiology**, v. 6, p.1-22, 2015.
- CROFTS, T. S.; GASPARRINI, A. J.; DANTAS, G. Next-generation approaches to understand and combat the antibiotic resistome. **Nature Reviews Microbiology**, v. 15, n. 7, p. 422–434, 10 abr. 2017.
- DUBILIER, N; MCFALL-NGAI; ZHAO, L. Create a global microbiome effort. **Nature**, v. 526, p.631-634, 2015.
- FLORENCIO, C.; BADINO, A. C.; FARINAS, C. S. Current challenges on the production and use of cellulolytic enzymes in the hydrolysis of lignocellulosic biomass. **Química Nova**, p.1082-1093, 2017.

- FUJITA, K. et al. Molecular Cloning and Characterization of a β -l-Arabinobiosidase in *Bifidobacterium longum* That Belongs to a Novel Glycoside Hydrolase Family. **Journal Of Biological Chemistry**, v. 286, n. 7, p.5143-5150, 2011.
- GONG, X. et al. Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. **Bmc Research Notes**, v. 5, p.566-578, 2012.
- HANDELSMAN, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. **Microbiology And Molecular Biology Reviews**, v. 68, n. 4, p.669-685, 2004.
- HUERTA-CEPAS, J. et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, v. 44, n. D1, p. D286–D293, 2016.
- HUGENHOLTZ, P.; HUBER, T. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. **International Journal Of Systematic And Evolutionary Microbiology**, v. 53, n. 1, p.289-293, 2003.
- HYATT, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, v. 11, p. 119, 2010.
- KATOH, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p.3059-3066, 15 jul. 2002.
- KATOH, K.; STANDLEY, D. M.. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. **Molecular Biology And Evolution**, v. 30, n. 4, p.772-780, 2013.
- KIM, M. et al. Analytical Tools and Databases for Metagenomics in the Next-Generation Sequencing Era. **Genomics & Informatics**, v. 11, n. 3, p.102-113, 2013.
- KISHI, L. T. et al. Draft Genome Sequence of a Chitinophaga Strain Isolated from a Lignocellulose Biomass-Degrading Consortium. **Genome Announcements**, v. 5, n. 3, p. e01056-16, 2017.
- KUMAR, S. et al. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. **Bioinformatics**, [s.l.], v. 28, n. 20, p.2685-2686, 2012.

KUNIN, V. et al. A Bioinformatician's Guide to Metagenomics. **Microbiology And Molecular Biology Reviews**, v. 72, n. 4, p.557-578, 2008.

LEARY, N. A. O. et al. Reference sequence (RefSeq) database at NCBI : current status , taxonomic expansion , and functional annotation. v. 44, p. 733–745, 2016.

MADHAVAN, A. et al. Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. **Applied Biochemistry And Biotechnology**, v. 183, n. 2, p.636-651, 2017.

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends In Genetics**, v. 24, n. 3, p.133-141, 2008.

MENEZES, C. R; BARRETO, A. R. Biodegradation of lignocellulosic wastes by basidiomycetes fungi: Characterization of waste and study of fungal enzyme complex. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v. 19, n. 2, p.1365-1391, 2015.

MOORE, A. D. et al. DoMosaics: software for domain arrangement visualization and domain-centric analysis of proteins. **Bioinformatics**, v. 30, n. 2, p.282-283, 2013.

NAVAS-MOLINA, J. A. et al. The microbiome and big data. **Current Opinion In Systems Biology**, v. 4, p.92-96, 2017.

NAVES, B. T. O. Panorama internacional das patentes biotecnológicas. **Revista de Informação Legislativa**, v. 197, n. 50, p.75-83, 2013.

NGARA, T. R.; ZHANG, H. Recent Advances in Function-based Metagenomic Screening. **Genomics, Proteomics & Bioinformatics**, v. 16, n. 6, p.405-415, 2018.

OBERTO, J. SyntTax: a web server linking synteny to prokaryotic taxonomy. **Bmc Bioinformatics**, v. 14, n. 1, p.1-10, 2013.

PEREIRA, G. L et al. Status of the art sequencing genome on livestock. **Ars Veterinaria**, v. 29, n. 3, p.190-199, 2013.

QUERO, S. et al. Proteómica en enfermedades infecciosas. **Enfermedades Infecciosas y Microbiología Clínica**, v. 34, n. 4, p.253-260, 2016.

QUIROGA, C. Las tecnologías «ómicas»: situación actual y desafíos futuros. **Revista Argentina de Microbiología**, v. 48, n. 4, p.265-266, 2016.

RIO, T. G. D et al. Complete genome sequence of Chitinophaga pinensis type strain (UQM 2034T). **Standards In Genomic Sciences**, v. 2, n. 1, p.87-95, 2010.

- SCHAFFERT, L et al. Complete genome sequence of the actinomycete *Actinoalloteichus hymeniacidonis* type strain HPA 177T isolated from a marine sponge. **Standards In Genomic Sciences**, v. 11, n. 1, p.1-8, 2016.
- SCHLOSS, P. D; HANDELSMAN, J. Biotechnological prospects from metagenomics. **Current Opinion In Biotechnology**, v. 14, n. 3, p.303-310, 2003.
- SHARPTON, T. J. An introduction to the analysis of shotgun metagenomic data. **Frontiers in Plant Science**, v. 5, 16 jun. 2014.
- SURETTE, M. D.; WRIGHT, G. D. Lessons from the Environmental Antibiotic Resistome. **Annual Review of Microbiology**, v. 71, n. 1, p. 309–329, 8 set. 2017.
- TATUSOV, R. L. et al. The COG database: an updated version includes eukaryotes. **BMC bioinformatics**, v. 4, p. 41, 2003.
- WHITE III, R. A. et al. ATLAS (Automatic Tool for Local Assembly Structures) -a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. p. 1–11, 2017.
- WILKENS, C. et al. GH62 arabinofuranosidases: Structure, function and applications. **Biotechnology Advances**, v. 35, n. 6, p.792-804, 2017.
- WOOLEY, J. C.; GODZIK, A.; FRIEDBERG, I. A Primer on Metagenomics. **Plos Computational Biology**, v. 6, n. 2, p.1-13, 2010.
- XING, M.; ZHANG, X.; HUANG, H. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. **Biotechnology Advances**, v. 30, n. 4, p.920-929, 2012.
- YIN, Y. et al. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. **Nucleic Acids Research**, v. 40, n. W1, p. 445–451, 2012.
- ZDOBNOV, E. M.; APWEILER, R. InterProScan - an integration platform for the signature-recognition methods in InterPro. **Bioinformatics**, v. 17, n. 9, p. 847– 848, 2001.

Capítulo II: Clonagem e expressão da sequência CB10_2.116.

1. INTRODUÇÃO

Nos últimos anos, graças ao desenvolvimento econômico de muitos países, ocorreu um crescimento no interesse por novos recursos provenientes de energia renovável, principalmente devido à crescente preocupação com as mudanças climáticas. Com isso houve um aumento nos estudos voltados ao uso de biomassa lignocelulósica como matéria-prima para o uso na produção de biocombustíveis (CARVALHO et al., 2018).

A lignocelulose, um dos principais compostos da biomassa vegetal, é composta principalmente por três polímeros: celulose, hemicelulose e lignina, dispostos em uma matriz tridimensional complexa dentro da qual os polissacarídeos são protegidos contra a degradação química e biológica dos agentes ambientais. As xilanas são a principal forma de hemicelulose encontrada em resíduos agroindustriais; consistem de um esqueleto de resíduos de β -1,4-D-xilopiranosose que podem ser ramificados, mostrando diferentes substituintes nas cadeias laterais, como a L-arabinofuranose, ácido glicurônico e/ou ácido 4-O-metilglicurônico (CARVALHO et al., 2018). Isso faz desse polímero um composto complexo que requer um conjunto de enzimas para sua completa desconstrução.

A degradação da biomassa vegetal tem se destacado por ser um importante campo de estudo para a aplicação de enzimas microbianas, como por exemplo para a produção de etanol lignocelulósico ou de segunda geração (etanol 2G). Tal processo consiste em utilizar enzimas ativas em carboidratos (CAZymes) para a hidrólise desses polissacarídeos. Essas enzimas são categorizadas em diferentes famílias de acordo com sua função no banco de dados do CAZy (<http://www.cazy.org>), incluindo glicosil hidrolases (GHs) que hidrolisam as ligações glicosídicas de carboidratos complexos, glicosil transferases (GTs) que transferem moléculas de açúcar, polissacarídeo liases (PLs) que atuam em cadeias de polissacarídeos contendo ácido urônico, carboidrato esterases (CEs) que hidrolisam as ligações éster em carboidratos e, finalmente, enzimas com atividades auxiliares (AAs) que promovem reações de oxidação-redução para degradação da biomassa vegetal (CONTESINI et al., 2017).

As arabinofuranosidases (Abfs), EC 3.2.1.55, têm recebido atenção especial devido à sua gama de aplicações, sendo conhecidas como enzimas auxiliares capazes de desfazer ligações α -L-arabinofuranosídicas, além de agir em conjunto com outras enzimas para hidrólise completa de hemicelulose. Vale ressaltar que tais enzimas representam uma parcela potencial de enzimas limitantes que atuam na degradação do material lignocelulósico (KURNIATI; DARMOKOESOEMO; PUSPANINGSIH, 2016)

A ação de α -L-arabinofuranosidases ou a associação com outras enzimas que degradam a hemicelulose representa uma ferramenta biotecnológica promissora como alternativa para algumas tecnologias químicas existentes, como por exemplo sua utilização na produção de compostos medicinais tais como agentes antiglicêmicos e anticarcinogênicos, melhoramento do sabor do vinho, aumento da qualidade do pão, tratamento de polpas vegetais, clarificação de sucos, aumento da qualidade de rações, produção de etanol 2G e síntese de oligossacarídeos (NUMAN; BHOSLE, 2005).

O Laboratório de Bioquímica de Microrganismos e Plantas (LBMP) da UNESP/FCAV de Jaboticabal descreveu o genoma de uma bactéria *Chitinophaga* sp. (Cadastro sisgen: A5F83B5) (KISHI et al., 2017). Este genoma derivou de estudos preliminares de um consórcio microbiano degradador de biomassa para produção de etanol lignocelulósico e teve o seu DNA total extraído e sequenciado. O genoma parcial obtido foi comparado ao genoma de uma *Chitinophaga eiseniae* sp. nov. (YASIR et al., 2011) proveniente dos bancos de dados do NCBI (*National Center for Biotechnology Information*).

O genoma de *Chitinophaga* sp. codifica enzimas capazes de degradar biomassa sendo utilizada para a prospecção de genes envolvidos na desconstrução de lignocelulose. Considerando os resultados obtidos por similaridade em procura na base de dados PDB (*Protein Data Bank*), UniProt (*Universal Protein Resource*) e PFAM (*Protein Families*) com melhores resultados de similaridade para uma arabinofuranosidase de *Chitinophaga*, o conjunto de resultados sugere que se trata de uma enzima ainda não descrita. O estudo de enzimas como a α -L-arabinofuranosidase, se mostra importante tanto do ponto de vista industrial quanto farmacêutico, podendo esse conferir benefícios no âmbito econômico.

Nesta etapa do trabalho o objetivo foi caracterizar estrutural e funcionalmente uma α -L-Afase codificada por uma ORF proveniente da mineração de dados metagenômicos, bem como a aplicação desta enzima para os açúcares residuais da produção de etanol.

2. REVISÃO DA LITERATURA

2.1 Biomassa lignocelulósica

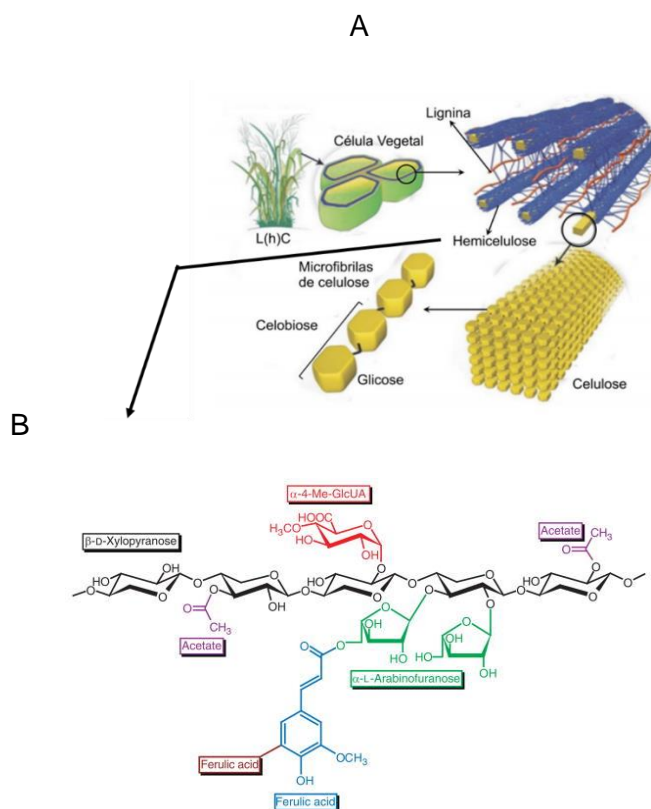
Nas últimas décadas é evidente o aumento de pesquisas com foco no desenvolvimento de tecnologias para a produção de biocombustíveis e outros produtos de interesse comercial a partir de biomassa vegetal. Isso se justifica em consequência de que a maioria do carbono fixado fotossinteticamente é incorporada aos polímeros da parede celular vegetal, tornando esse material uma fonte abundante de biomassa terrestre (CARVALHO; BORTOLINI; BARCELLOS, 2015).

A produção de energia renovável a partir da biomassa vegetal também atende a vários requisitos de sustentabilidade, sendo uma aposta interessante para substituir, ainda que parcialmente, os combustíveis fósseis. O uso potencial da biomassa vegetal na produção de biocombustíveis, especialmente o etanol de segunda geração, levou a um aumento nos estudos com o objetivo de alcançar uma desconstrução eficiente da lignocelulose. As enzimas lignocelulolíticas são avaliadas através do isolamento de microrganismos produtores, da sua análise bioquímica e da prospecção de genes que codificam enzimas, tanto dos diferentes microrganismos como em metagenomas (CAMARGO et al., 2018; SANTOS et al., 2012).

A parede celular vegetal é composta por uma mistura de polissacarídeos, proteínas, compostos fenólicos e sais minerais. A lignocelulose representa mais de 90% do peso seco de uma célula vegetal, sendo composta pelos polímeros celulose, hemicelulose e lignina, unidos fortemente entre si por forças não covalentes e ligações covalentes (figura 1). Nas paredes celulares de tecidos vasculares de plantas terrestres superiores, as fibrilas de celulose estão localizadas

em uma matriz amorfa de lignina e hemicelulose. A quantidade de cada um dos polímeros varia com a espécie e a idade da planta, bem como entre as partes de uma planta e de acordo com o clima predominante. Em média, a lignocelulose consiste de 45% de celulose, 30% de hemicelulose e 25% de lignina. Essa matriz é altamente ordenada e dinâmica podendo tornar-se mais rígida ou mais frouxa conforme as necessidades da planta (MENEZES; BARRETO, 2015; CONTESINI et al., 2017).

Figura 1 - (A) Componentes estruturais da parede vegetal e estrutura geral mostrando (B) as várias ligações encontradas em xilanas isolados das paredes celulares da planta.

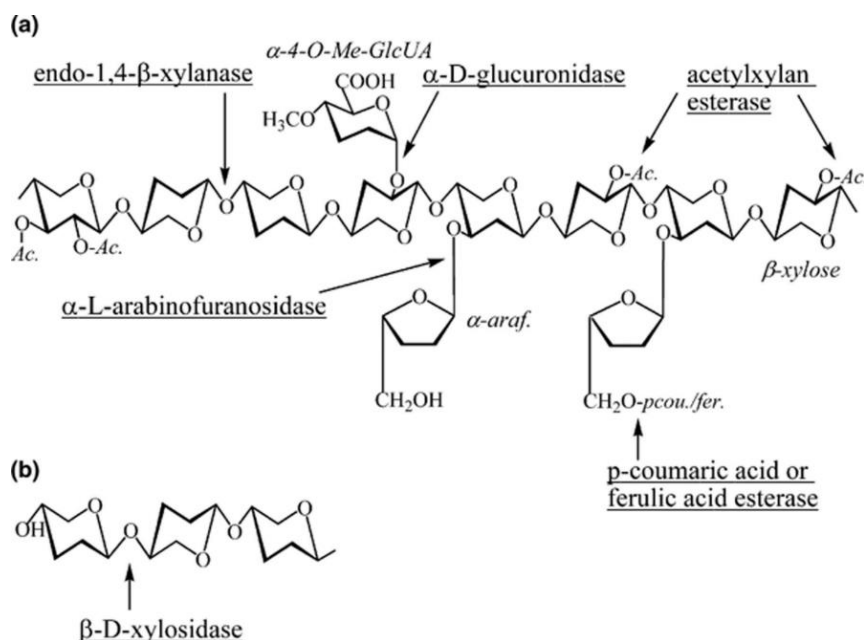


Fonte: Adaptado de Santos et al, 2012.

Algumas dificuldades são encontradas no processo de conversão da biomassa lignocelulósica em insumos químicos, essas são atribuídas às características químicas e morfológicas desse material, uma vez que, sua constituição de fibras de celulose envolvidas em uma matriz amorfa de polioses e lignina, age como uma barreira natural ao ataque de microrganismos e/ou enzimas tornando esse material estruturalmente rígidos e pouco reativos (SANTOS et al., 2012).

A celulose é o homopolímero natural de maior ocorrência no mundo, caracterizada por sequências de resíduos β -D-glicopiranosídicos anidro unidos covalentemente. Já as hemiceluloses são heteropolissacarídeos complexos compostos por D-glicose, D-galactose, D-manose, D-xilose, L-arabinose, ácido D-glicurônico e ácido 4-O-metil-glicurônico (figura 2), sua estrutura apresenta ramificações que interagem facilmente com a celulose, dando estabilidade e flexibilidade ao agregado (SANTOS et al., 2012).

Figura 2 – A. Estrutura de xilana e os locais de clivagem por enzimas xilanolíticas. A espinha dorsal do substrato é composta por resíduos de xilose ligados por ligação β -1,4. As arabinoxilanas encontradas nas paredes celulares das plantas de cereais e gramíneas, contêm uma espinha dorsal de xilana parcialmente substituída em intervalos com resíduos de α -L-arabinofuranose. B. Local de clivagem da enzima xilosidase.



Fonte: Collins; Gerday; Feller (2005)

A biomassa vegetal, em geral, é composta principalmente de celulose e hemicelulose, seguida de lignina. Sendo assim, cabe ressaltar que o bagaço de cana-de-açúcar se apresenta como uma das fontes mais importantes de resíduos lignocelulósicos, sendo composto de 25-45% de celulose, 28-32% de hemicelulose e 15-25% de lignina. A principal hemicelulose é a arabinoxilana, constituída por um esqueleto de xilose com ligações β -(1-4) ramificadas com unidades de arabinofuranosila e 4-O-metilglucopiranosila (CONTESINI et al., 2017).

2.2 Hemicelulose

As hemiceluloses incluem uma série de polissacarídeos dos quais, podemos encontrar as xilanas como o segundo mais abundante na natureza. A hemicelulose é um polissacarídeo heterogêneo encontrado nas plantas e na parede celular de alguns microrganismos. A completa degradação de polímeros de xilana como componentes lignocelulósicos da parede celular requer um conjunto de enzimas hidrolíticas (JIA et al., 2015).

As xilanas são um grupo de polissacarídeos com uma estrutura central composta por monómeros de xilose unidos por ligações β -1,4. O esqueleto de xilana pode ser ramificado com uma ampla gama de açúcares e outros grupos laterais, tais como xilose, arabinose, galactose e ácido glicurônico (MECHELKE et al., 2017).

Cadeias de xilana podem ser amplamente classificadas como homoxilanas, arabinoxilanas, glucuronoxilanas e arabinoglucuronoxilanas, sendo as arabinoxilanas as principais formas de hemicelulose nas paredes celulares da planta e consistem em um esqueleto de xilose com resíduos de arabinose ligados ao seu O-2 ou O-3 (GONÇALVES et al., 2015).

A degradação do esqueleto de xilana por xilanases tem se mostrado pouco eficiente devido à presença de ramificações de arabinose; em virtude disso se torna importante a remoção dessas cadeias laterais pela adição de arabinofuranosidase, com a finalidade de melhorar a eficiência na degradação da hemicelulose. A ação cooperativa entre atividades de arabinofuranosidases e xilanases pode ser utilizada para melhorar a liberação de sacarídeos solúveis durante a hidrólise de arabinoxilana. A função da arabinofuranosidase na desconstrução de xilana é

remover cadeias laterais de arabinose do esqueleto de xilose de arabinoglucuronoxilana. Nos últimos anos, a biodegradação de hemicelulose atraiu atenção na indústria, uma vez que é possível realizar a conversão de biomassa em açúcares fermentáveis para o uso na produção de biocombustíveis e produtos químicos de valor agregado (JIA et al., 2015).

2.3 Famílias de glicosil hidrolases

Os carboidratos complexos são amplamente distribuídos na natureza, e estão envolvidos em uma infinidade de funções biológicas, da reserva de carbono a moléculas estruturais. A diversidade de carboidratos complexos é controlada por um conjunto de enzimas envolvidas na sua montagem (glicosiltransferases) e sua degradação (glicosídeos hidrolases, polissacarídeos liases, carboidrato esterases), coletivamente designados como enzimas ativas de carboidratos (CAZymes). Avanços sustentáveis na degradação da parede celular das plantas dependem profundamente de enzimas ativas em carboidratos. Para várias dessas enzimas existe ainda a necessidade de atingirmos uma melhor compreensão das relações estrutura/função, bem como de descobrirmos fontes biológicas de enzimas com potencial significativo de aplicação (WILKENS et al., 2017; LOMBARD et al., 2013).

A desconstrução do material lignocelulósico pela ação de enzimas ocorre graças a um complexo de enzimas, as quais podem ser produzidas por diversos microrganismos que em conjunto trabalham para clivar as macromoléculas (polissacarídeos) em monômeros ou oligômeros de açúcares, fornecendo fonte de carbono e energia para os microrganismos. A cooperação da família de glicosil hidrolases, incluindo a endoxilanase (EC 3.2.1.8) e a β -xilosidase (EC 3.2.1.37), assim como enzimas acessórias, entre elas α -L-arabinofuranosidase, α -glicuronidase resulta na biodegradação completa de tais polímeros (MOHSEN DEHNAVI; PAZUKI; VOSSOUGH, 2015).

Enzimas ativas sobre carboidratos logo no início foram classificadas de acordo com o mecanismo de reação e especificidade do substrato e receberam um número de EC pela Comissão de Enzimas da União Internacional de Bioquímica e Biologia Molecular (IUBMB). No entanto, nos anos 90, o banco de dados

*Carbohydrate Active EnZyme (CAZy)*² começou a categorizar as sequências primárias dessas enzimas, que atualmente resultaram em 153 famílias de glicosil hidrolases (GH), e dentro de cada uma delas os membros compartilham estruturas e mecanismos catalíticos. Cabe aqui destacar que, as enzimas de uma dada família de GH podem ter diferentes especificidades de substrato, a mesma especificidade de substrato pode ser representada em diferentes famílias de GH (WILKENS et al., 2017).

Enzimas como as arabinofuranosidases são encontradas em diferentes famílias de GHs e podem hidrolisar ligações glicosídicas com inversão líquida (GH 43) ou retenção (GH 51, 54) da configuração estereoquímica no carbono anomérico (GONÇALVES et al., 2015). Entre essas famílias, as GH 43 tem uma característica distinta. O sítio ativo das β -xilosidases e arabinofuranosidases é altamente conservado, consistente com a observação de que as β -xilosidases e algumas arabinofuranosidases são capazes de clivar as ligações de α -arabinofuranosila e β -xilopiranosila, respectivamente (MOHSEN DEHNAVI; PAZUKI; VOSSOUGH, 2015).

As enzimas pertencentes à família GH43 incluem as β -xilosidases (EC 3.2.1.37), α -L-arabinofuranosidases (EC 3.2.1.55), dentre outras. Todas as famílias GH realizam hidrólise com retenção da configuração anomérica, exceto a GH 43, que é inversora da mesma. Algumas enzimas GH43 mostram uma função dupla da β -D-xilosidase e α -L-arabinofuranosidase devido à semelhança conformacional entre xilopiranosose e arabinofuranose, permitindo assim essa propriedade (CINTRA et al., 2017).

2.4 Aplicações biotecnológicas de arabinofuranosidases

As ABFs são enzimas que apresentam propriedades bioquímicas diversas, permitindo utilização em vários ramos da biotecnologia. A crescente demanda por processos mais eficientes e menos poluidores resulta em um aumento por reagentes sustentáveis, com baixo impacto ambiental e que atuem com maior

² www.cazy.org

rendimento. As hemiceluloses começaram a ser utilizadas há mais de 50 anos na indústria de alimentos, seguidos da indústria têxtil e de papel e celulose. Atualmente, elas apresentam diversos ramos de aplicação e, especialmente as ABFs podem ser aplicadas na agricultura, na degradação de materiais hemicelulósicos, no desenvolvimento de fármacos, na indústria de alimentos e na nutrição, na produção de polpa de papel e ainda na medicina. É importante salientar que em muitas aplicações, as ABFs, por serem enzimas acessórias, são importantes pela sua ação sinérgica com outras enzimas do complexo xilanolítico, pois muitas vezes enzimas com ação endo – como as xilanases – ou com ação exo – como as β -xilosidases – não conseguem acessar o seu substrato, devido a presença de ramificações de arabinose (NUMAN; BHOSLE, 2005).

As ABFs são importantes na degradação eficiente e completa das hemiceluloses e pectinas da parede celular de plantas, e sua aplicação em processos industriais tem benefícios econômicos e ambientais relevantes, como exemplo o uso em biorrefinarias para a produção de bioetanol de segunda geração. Uma solução alternativa promissora consiste na degradação enzimática das hemiceluloses utilizando um coquetel enzimático composto de arabinofuranosidases, xilanases, esterases e xilosidases, que promoveriam uma eficiente conversão da hemicelulose em monossacarídeos. A utilização de polímeros como a xilana é importante para o desenvolvimento de sistemas de biorrefinaria viáveis economicamente, e as estratégias de valorização da xilana têm por objetivo principalmente sua hidrólise total a açúcares fermentáveis como substratos para processos de fermentação (MECHELKE et al., 2017; WILKENS et al., 2017).

Na indústria de alimentos, as ABFs são bastante utilizadas no processamento de frutas, vegetais e cereais, pois, juntamente com outras enzimas hemicelulolíticas, como as xilanases, arabinanases e as pectinases, atuam na quebra dos polissacarídeos presentes no filtrado e no suco de maceração. Outra aplicação das ABFs dentro desse setor é na indústria de vinhos, atuando no processo de fabricação da bebida pois auxiliam na liberação de terpenóis – compostos aromáticos responsáveis pela aromatização dos vinhos. Esses compostos se localizam na casca da uva onde estão de forma livre, aromática, ou ligados à uma molécula de açúcar (forma não-aromática). Geralmente esse açúcar

é uma arabinose, o que confere às ABFs a função de hidrolisar a ligação, liberando o terpeno para sua forma aromática, o que por sua vez, irá aumentar o aroma do vinho, aumentando assim sua qualidade de mercado (NUMAN e BHOSLE, 2005).

Outra aplicação das ABFs no ramo alimentício é na fabricação de ração animal. Como a base da alimentação de frangos e porcos é a ração, composta principalmente por material lignocelulósico, estes animais apresentam um tempo de digestão mais prolongado. Com a utilização de celulasas, pectinases e hemicelulasas muitas vezes a solubilização não é suficiente, pois as enzimas glicosídicas não conseguem acessar as regiões da cadeia ramificada. Com a adição de ABFs, que atuarão sobre essas ramificações, as enzimas glicosídicas acessam as ligações alvo e promovem a solubilização do material, facilitando assim a digestão e o maior aproveitamento do alimento pelos animais (SHALLOM et al., 2002).

ABFs podem ser aplicadas também na produção de xilo-oligossacarídeos (XOS). Esses são oligômeros com baixo grau de polimerização, formados por cadeias curtas de xilose, não ramificadas. Eles podem ser utilizados como prebióticos, substâncias que favorecem o crescimento de bactérias da flora intestinal humana – os probióticos. Os XOS podem ser produzidos quimicamente, mas o uso de enzimas xilanolíticas é mais eficaz, pois elas atacam diretamente as ligações internas da cadeia polissacarídica, quebrando o polímero em cadeias menores, de 2 a 5 xiloses. Nesse contexto, as ABFs atuam em cooperação com as xilanases, hidrolisando as ramificações de arabinose, preparando o açúcar para a atuação das xilanases (BRIENZO; CARVALHO; MILAGRES, 2010; CHAPLA; PANDIT; SHAH, 2012).

A arabinose pode ser utilizada como um adoçante de baixo teor calórico, devido ao sabor doce que apresenta, com baixa absorção pelo organismo humano. Além disso, ela inibe seletivamente a absorção de sacarose, de maneira não competitiva, quando ingerida ao mesmo tempo, o que permite a redução do índice glicêmico, prevenindo e tratando a hiperglicemia após as refeições em pacientes diabéticos e inibindo a obesidade (NUMAN; BHOSLE, 2005).

Na indústria de polpa de papel, as ABFs podem ser aplicadas aumentando a eficiência das soluções de xilanases, facilitando a deslignificação da polpa. As ABFs atuam sobre as hemiceluloses, liberando os resíduos de arabinose que

muitas vezes impedem a ação das enzimas do branqueamento, o que permite a remoção da lignina ligada às hemiceluloses. Para que essas enzimas possam atuar de maneira eficiente nesse processo, é necessário que elas apresentem alta estabilidade a temperaturas elevadas, em pH alcalino e ausência de atividade celulolítica (NUMAN; BHOSLE, 2005).

Como já visto anteriormente, a matriz lignocelulósica é bastante complexa, sendo necessário a atuação de diversas enzimas em conjunto para a eficiente conversão dos polissacarídeos em açúcares fermentescíveis. Assim, as ABFs com capacidade de degradar biomassa, se tornam fundamentais para o processo de sacarificação das paredes celulares e fibras de resíduos florestais e agrícolas (YANG et al., 2015).

2.5 Expressão heteróloga de proteínas em *Escherichia coli*.

Certamente a produção de proteínas recombinantes em sistemas microbianos revolucionou a Bioquímica. Anteriormente grandes quantidades de tecidos de animais e plantas ou grandes volumes de fluidos biológicos eram necessários para a purificação de pequenas quantidades de uma determinada proteína. Atualmente pesquisadores que desejam iniciar um projeto que necessite de uma proteína purificada, podem obtê-la de forma recombinante, graças à possibilidade de expressar e purificar a proteína heteróloga em grande quantidade, o que permite sua caracterização bioquímica, seu uso em processos industriais e o desenvolvimento de bens comerciais (ROSANO; CECCARELLI, 2014).

Existem muitos hospedeiros utilizados para a produção de proteínas recombinantes, tais como bactérias, leveduras, insetos, plantas e animais. A bactéria *E. coli* tem sido o meio mais popular de produzir proteínas recombinantes por mais de duas décadas, não apenas de origem procariótica, mas também eucariótica, pois pode crescer em meios de baixo custo sob condições laboratoriais bem definidas. Devido ao seu rápido crescimento, a seleção de mutantes é fácil e conveniente. Além disso, as células de *E. coli* são altamente eficientes para receber (incorporar) DNA de outro organismo e expressar proteínas recombinantes a uma elevada taxa (KAUR; KUMAR; KAUR, 2018).

Teoricamente, os passos necessários para obter uma proteína recombinante são bem diretos. Quando se tem o gene de interesse descrito, ele pode ser clonado em qualquer vetor de expressão que atenda às necessidades desejadas, e transformado no hospedeiro desejado, induzido e, então, a proteína estaria pronta para purificação e caracterização. Na prática, no entanto, são encontradas dificuldades. Por exemplo, o fraco crescimento do hospedeiro, a formação de corpos de inclusão, a inatividade da proteína e até mesmo a não obtenção da proteína heteróloga são alguns dos problemas frequentemente encontrados nesse processo (ROSANO; CECCARELLI, 2014).

Os principais componentes de um vetor típico de plasmídeo com influência na eficiência da transcrição incluem origem de replicação (Ori), promotor, marcador de seleção de antibiótico e terminadores de transcrição (SINGHA et al., 2017).

A escolha da estirpe hospedeira de *E. coli* apropriada é obrigatória para a expressão substancial de proteínas recombinantes e depende do vetor de expressão utilizado. Para a expressão do gene no vetor contendo promotores T7, uma estirpe de *E. coli* BL21 é preferida porque produz a RNA polimerase para este promotor. BL21 é uma cepa robusta de *E. coli*, capaz de crescer vigorosamente em meio mínimo. As bactérias BL21 (DE3) são convenientes para a expressão de proteínas utilizando o promotor T7. Os promotores *lac*, *tac* e *trc* são reconhecidos pela RNA polimerase de *E. coli* e, portanto, podem ser expressos em qualquer *E. coli*, como BL21, Top10 ou DH5 α (KAUR; KUMAR; KAUR, 2018).

Diversos sistemas de expressão estão disponíveis para várias aplicações. Aproximadamente 80% das proteínas com estruturas tridimensionais resolvidas submetidas ao banco de dados de proteínas (PDB) em 2003 foram expressas em *E. coli*. Um vetor de expressão é projetado de forma que um gene clonado seja transcrito e/ou traduzido. Os vetores plasmídicos são utilizados para a clonagem e expressão de genes estranhos no sistema procariótico. O vetor de expressão usado, portanto, possui elementos específicos para uso no sistema particular (KAUR; KUMAR; KAUR, 2018).

As vantagens da utilização de *E. coli* como organismo hospedeiro, além das citadas acima, incluem: (i) crescimento rápido; em meios de sais nutrientes e glicose e dadas as condições ambientais ótimas, seu tempo de duplicação é de cerca de 20 min.; (ii) culturas de alta densidade celular são facilmente

alcançadas; (iii) meios complexos ricos podem ser feitos de componentes disponíveis e de baixo custo; (iv) a transformação com DNA exógeno é rápida e relativamente simples (ROSANO; CECCARELLI, 2014).

Proteínas com menos de 100 kDa são bem toleradas em *E. coli* e, portanto, são expressas em níveis significativos. Também foi sugerido que o tamanho da proteína pode afetar a eficiência da secreção e grandes proteínas citoplasmáticas são fisicamente impossíveis de translocar. Se o tamanho exceder este limite, então sistemas hospedeiros diferentes de *E. coli* devem ser usados, por exemplo, expressão em células de insetos. Por outro lado, proteínas pequenas e polipeptídeos (<10 kDa) são difíceis de expressar de forma estável em *E. coli* por causa do seu envelhecimento inadequado e são frequentemente submetidos a degradação proteolítica. Estas proteínas podem ser estabilizadas, expressando-as em fusão a uma proteína grande, que lhes permite dobrar adequadamente (KAUR; KUMAR; KAUR, 2018).

A indução da expressão de proteínas é conhecida como o mecanismo pelo qual a transcrição de genes estruturais de um operon aumenta em resposta à presença de um substrato específico no meio. Isso se deve pelo fato da síntese de uma proteína requerer grandes quantidades de energia e recursos, tornando inviável para a célula a expressão de todos os genes presentes em seu DNA. Na regulação gênica em *E. coli* um gene indutível é aquele cuja transcrição é ativada pela adição de um indutor ao meio (ROZENDO et al., 2014).

Os indutores se ligam a sítios específicos que podem estar localizados nas proximidades ao promotor, resultando no favorecimento da transcrição pelo aumento da afinidade da RNA-polimerase. Assim, tal composto químico também apresenta a capacidade de regular a expressão do gene. Isso pode ser uma vantagem para a produção de proteínas recombinantes uma vez que é possível o controle de sua síntese. Com base no conhecimento detalhado sobre o operon lac de *E. coli* e seu fácil controle por indutores químicos, os sistemas de expressão em bactérias mais comumente utilizados são baseados nesse sistema (ROZENDO et al., 2014).

O indutor natural do operon lac é a lactose, mas existem outros indutores utilizados para esses sistemas de expressão, sendo o mais comum o isopropil- β -D-tiogalactopiranosídeo (IPTG). O IPTG é um dos indutores dos sistemas de

expressão baseados no operon lac. Esse indutor liga-se à proteína repressora, LacI, levando à diminuição de sua afinidade pela região promotora do operon lac. O IPTG representa hoje um dos indutores mais utilizados em sistemas de alta expressão, no entanto uma das grandes dificuldades na indução por IPTG é sua alta citotoxicidade, que dependendo da concentração pode produzir efeitos negativos no crescimento celular (BRIAND et al., 2018).

2.6 Análises *In silico*

A biotecnologia se utiliza de conhecimentos sobre os processos biológicos e sobre as propriedades dos seres vivos, com o fim de resolver problemas e criar produtos de utilidade para diferentes finalidades. Esse processo surgiu da necessidade de se suprir as transformações globais que ocorreram na ciência e no mercado. A biotecnologia está intimamente relacionada à inovação tecnológica, pois propõe o desenvolvimento de novas tecnologias e produtos, aplicando as informações desenvolvidas na pesquisa. Os estudos com biotecnologia envolvem pesquisas sobre transgênicos, genômica, proteômica, terapia gênica, entre outras (SANTOS FILHO; OSVALDO; ALENCASTRO, 2003).

Dentro do campo da biotecnologia se pode destacar a bioinformática como meio para o desenvolvimento de métodos e ferramentas de software para entender dados dos mecanismos biológicos. Esse campo combina ciência da computação, biologia, matemática e técnicas estatísticas para analisar e interpretar dados biológicos. A bioinformática além de outras ferramentas possibilita a pesquisa em bancos de dados, com a finalidade de analisar dados e simular experimentos, onde propõe novas formas de ciência baseada na experimentação *in silico*, prevendo estruturas de proteínas e moléculas, realizar testes de interação, inibição ou ativação de moléculas, criar inibidores, moléculas de interferência, entre outras atividades. O desenvolvimento da bioinformática está relacionado à biotecnologia a partir do momento que geramos novos dados e conhecimentos que podem ser aplicados para o desenvolvimento de novos produtos e soluções (SOUZA; RHODEN; PAMPFILE, 2014).

O domínio da bioinformática cresceu em paralelo e com a internet, em que a rápida análise de dados e a troca de informações sobre os códigos biológicos e computacionais estão em convivência harmônica, por meio de suas múltiplas ramificações, gerenciando e integrando bancos de dados, e construindo sistemas *in silico* para simulação de formas naturais e modificadas de produtos específicos (ESPINDOLA et al., 2010).

No estudo completo das proteínas, integrando estrutura e função, os pesquisadores utilizam bancos de dados diversos que possam atender os diferentes ramos da proteômica. Um dos mais usados é o banco de dados UniProt, um catálogo de dados de sequências e funções de proteínas, mantido pelo consórcio UniProt. A missão da UniProt é fornecer à comunidade científica um recurso abrangente, de alta qualidade e de livre acesso à sequência de proteínas e informações funcionais (<https://www.uniprot.org/>).

O grande desafio enfrentado por estudiosos e bioinformatas é descobrir qual a estrutura tridimensional adotada pelas proteínas a partir da estrutura primária. Um método alternativo e não-experimental é a modelagem molecular, baseada em conhecimentos estereoquímicos dos aminoácidos. Uma das maneiras de se fazer a modelagem molecular é através da homologia entre sequências, em que uma delas já possui forma tridimensional definida. Houve um crescimento explosivo nos dados estruturais disponíveis para proteínas, não apenas a partir de cristalografia de raios X, mas também de estudos de ressonância magnética e microscopia eletrônica, ao que se somam avanços significativos em técnicas e hardware computacionais (SAXENA et al., 2009).

O mecanismo de duplicação de genes, associado às mutações, leva a divergências moleculares e, conseqüentemente, à formação de famílias de proteínas estruturalmente relacionadas. Proteínas derivadas de um ancestral comum são ditas homólogas. Em função do número de mutações envolvidas, as sequências de aminoácidos de proteínas homólogas podem ser, idênticas, semelhantes ou diferentes. As estruturas tridimensionais de proteínas homólogas tendem a se conservar porque a estrutura ancestral comum é crucial para a manutenção da função das proteínas. A conservação de resíduos em proteínas homólogas é notável. Sequências de resíduos de aminoácidos de proteínas com cerca de 30% de identidade, apenas, podem ter excelente sobreposição das

cadeias principais (*protein backbone*) (SANTOS FILHO; OSVALDO; ALENCASTRO, 2003).

O primeiro passo é a pesquisa de proteínas homólogas em bancos de dados de estruturas de proteínas como o PDB (*Protein Data Bank*), que é uma colaboração entre o RCSB (*Research Collaboratory for Structural Bioinformatics*), e o MSD-EBI (*Macromolecular Structural Database*). A seguir, deve-se realizar o alinhamento das sequências de resíduos das proteínas homólogas e a proteína alvo, através de um programa de múltiplo alinhamento, Clustal, por exemplo. A modelagem é realizada através de softwares como o Modeller, SWISS-MODEL, 3D-PSSM, dentre outros. Esses programas normalmente procuram encontrar a estrutura terciária que melhor se aproxime da disposição dos átomos das proteínas utilizadas como modelo, e ao mesmo tempo atenda às restrições físico-químicas (ESPINDOLA et al., 2010).

Por fim, o Pfam é um banco de dados que apresenta uma grande coleção de famílias de proteínas, cada uma representada por múltiplos alinhamentos de sequências e modelos ocultos de Markov (HMMs). As proteínas geralmente são compostas de uma ou mais regiões funcionais e de enovelamento mais compacto, comumente denominadas domínios. Diferentes combinações de domínios dão origem à diversidade de proteínas encontradas na natureza. A identificação de domínios dentro das famílias de proteínas pode, portanto, fornecer informações importantes sobre sua função (CHEN; HUANG; WU, 2017).

Os dados apresentados para cada entrada baseiam-se nos Proteomas de Referência UniProt, mas a informação sobre sequências individuais de UniProtKB pode ainda ser encontrada inserindo o acesso às proteínas. Os alinhamentos completos do Pfam estão disponíveis a partir da pesquisa de vários bancos de dados. As entradas de Pfam podem ser classificadas como Família (regiões proteicas relacionadas), Domínio (unidade estrutural proteica), Repetição (múltiplas unidades estruturais proteicas curtas), Motivos (unidade estrutural proteica curta fora dos domínios globais (ESPINDOLA et al., 2010).

As entradas de Pfam relacionadas são agrupadas em clãs com base na similaridade de sequência, estrutura ou perfil-HMM. O site de banco de dados Pfam fornece interface de pesquisa para consulta por sequência, palavra-chave, arquitetura de domínio, taxonomia e interfaces de navegação para analisar

sequências de proteínas para correspondências Pfam e visualizar anotações Pfam em arquiteturas de domínio, alinhamentos de sequência, interações, espécies e estruturas de proteínas no PDB (CHEN; HUANG; WU, 2017).

3. OBJETIVOS

Caracterizar estrutural e funcionalmente uma alfa-L-arabinofuranosidase bacteriana codificada por uma ORF proveniente da mineração de dados metagenômicos, bem como a aplicação desta enzima para auxiliar na degradação de lignocelulose a fim de obter a liberação de açúcares fermentáveis para a produção de etanol de segunda geração. Neste contexto, os seguintes objetivos específicos são definidos:

- Construir modelo da estrutura tridimensional da proteína através da análise de bioinformática para a caracterização da enzima;
- Clonar e expressar o gene codificador da enzima em vetor de expressão pET28a;
- Efetuar a extração e purificação da enzima por meio de cromatografia de afinidade para cauda de histidina e exclusão molecular;
- Caracterizar a enzima por meio de ensaios dos parâmetros cinéticos, análise do efeito de diferentes pHs, temperatura, estudo da estabilidade, efeito de íons e de diversas substâncias químicas na atividade enzimática;

4. MATERIAL E MÉTODOS

4.1 Análise *in silico* das sequências

Sequências codificadoras da enzima α -L-arabinofuranosidases foram prospectadas por similaridade e analisadas *in silico*, a identificação dos domínios para arabinofuranosidase foi analisada no banco Pfam (FINN et al., 2016) e os dados físico-químicos teóricos gerados a partir do ProtParam ExPASy (<http://expasy.org/>). Foi, ainda, realizada a certificação de possíveis peptídeos sinais, através de uma análise com o servidor SignalP 4.1 (PETERSEN et al., 2011). A predição da estrutura proteica foi obtida por métodos de modelagem por homologia empregando o “web server” Swiss Model (<https://swissmodel.expasy.org/>), e as estruturas foram analisadas através do programa Pymol versão 1.8 (The PyMOL Molecular Graphics System, Schrödinger, LLC).

4.2 Clonagem da sequência codificadora de arabinofuranosidase

Após selecionada a sequência de interesse experimentos de clonagem foram realizados. Inicialmente à amplificação da ORF selecionada foi executada através de PCR com a enzima Pfu DNA Polimerase, uma vez que esta enzima apresenta menor taxa de erro conhecida nas ampliações. Após identificada a ORF codificadora de enzima da família GH 127 no genoma de *Chitinophaga* sp., procedeu-se com a amplificação da sequência escolhida para clonagem com vetor pET28a. Este vetor caracteriza-se por possuir uma sequência codificadora de uma cauda de poli histidina na região N-terminal da proteína super-expressa que permite a purificação da proteína por cromatografia de afinidade. Inicialmente foram desenhados oligonucleotídeos iniciadores (*primers*) com sítios de restrição adequados nas regiões inicial e final do gene para a clonagem direcional (tabela 1).

Tabela 1. Propriedades dos oligonucleotídeos iniciadores de síntese utilizados para amplificação do gene em estudo.

Sequência	Enzima de restrição	Oligonucleotídeos (sequência 5' – 3')	TM	Tamanho do fragmento
CB10_2.116	BamHI_F	CGTACA GGATCC T CCACAAGCAGCCGCTTTCC	62	2340 pb
	EcoRI_R	TACGCA GAATTC CCGTTTGC GTGATATCGCCC	59,1	

Fonte: Autoria própria.

As reações de PCR foram realizadas nas seguintes condições: 2µL de Tampão 10X [Tris-HCl 200mM, pH 8,8; 100mM (NH₄)₂SO₄; 100mM KCl; 1% (v/v) Triton X-100]; 0,5µL de dNTP (4mM); 0,5µL de MgCl₂ (50mM); 5pmol de cada primer; 0,6µL de Taq Polimerase; 50ng de DNA genômico de *Chitinophaga* sp.; e água ultrapura para totalizar o volume final de reação de 20µL.

Para amplificação do gene as reações de PCR foram submetidas ao seguinte programa do termociclador: passo inicial a 95°C por 2 minutos, 30 ciclos a 95°C por 45 segundos, 60°C por 45 segundos, 62°C por 2 minutos e 30 segundos, e um passo final de 72°C por 10 min e estocadas a 4°C.

Novas reações foram feitas utilizando-se uma enzima de alta fidelidade e baixa taxa de erro, a *Pfu* DNA Polimerase. As reações foram realizadas nas seguintes condições: 45 µL de Super Mix (Platunum Super Mix – High Fidelity - Invitrogen), 2 µL (10pmol) de cada primer (*forward* e *reverse*); 3µL de DNA (50ng/µL) para totalizar o volume final de reação de 50µL. As reações foram submetidas ao programa descrito abaixo para melhor ajuste.

Após o término da reação uma alíquota de 2µL das reações de PCR foi aplicada em gel de agarose 1% contendo brometo de etídeo (0,5mg/µL), a fim de se constatar a amplificação dos produtos de PCR. A imagem foi visualizada e documentada sob luz UV em aparelho fotodocumentador Gel Doc 1000 (Bio-Rad, Hercules, Califórnia, USA), através do software Quantity OneR (Bio-Rad).

4.3 Eletroforese preparativa

O produto da PCR foi submetido a eletroforese preparativa em gel de agarose 1% com baixo ponto de fusão e isento de brometo de etídeo. Esse procedimento tem como finalidade eliminar bandas inespecíficas presentes no gel de agarose. Dessa maneira, a banda de interesse foi cortada e retirada do gel. A eluição da banda foi feita utilizando kit *Zymoclean Gel DNA Recovery* em conformidade com as instruções do fabricante. O material foi quantificado em espectrofotômetro NanoDrop™ 1000 (Thermo Fisher Scientific, Wilmington, Delaware, USA), sendo que a qualidade do DNA e a presença de proteínas contaminantes foi verificada pela relação 260/280nm.

4.4 Digestão do vetor e inserto com BamHI

4.4.1 Digestão de Vetor

Para essa reação foram usados 2µL de Buffer 10x (Fast digest), 2µL da enzima BamHI, 7µL de DNA e 9µL de água ultrapura para completar um volume final de 20µL. A reação foi submetida ao seguinte programa no aparelho termociclador, 30 minutos a 37°C (Disable) e 5 minutos a 80°C (Enable). A purificação do material se deu com o kit DNA clean e concentrator.

4.4.2 Digestão do Inserto

Para essa reação foi usado 2µL de Buffer 10x (*Fast digest*), 2µL da enzima BamHI, 13,6µL de DNA e 2,4µL de água ultrapura para completar um volume final de 20µL. A reação foi submetida junto com a anterior ao seguinte programa no aparelho termociclador, durante 30 minutos a 37°C (*Disable*) e 5 minutos a 80°C (*Enable*). A purificação do material foi feita com o *DNA Clean & Concentrator Kit* (Zymo Research).

4.5 Digestão do vetor e inserto com EcoRI

4.5.1 Digestão de Vetor

O vetor pET28a (Novagen, Gibbstown, New Jersey, USA) foi escolhido, sendo que para essa reação foram usados 2µL de Buffer da EcoRI, 1µL da enzima EcoRI, 13µL de DNA e 4µL de água ultrapura para completar um volume final de 20µL. A reação foi submetida ao aparelho termociclador a 37°C por 15 horas (*overnight*).

4.5.2 Digestão de Inserto

Para essa reação se usou 2µL de Buffer da EcoRI, 1µL da enzima EcoRI, 13µL de DNA e 4µL de água ultrapura para completar um volume final de 20µL. A reação foi submetida ao seguinte programa no aparelho termociclador: 37°C por 15 horas (*overnight*). Após as 15 horas de reação, no vetor foi adicionado “*Shrimp Alkaline Phosphatase*” (SAP) (1 U/µL, Fermentas, USA), em uma concentração de 1,13pmol (<http://www.promega.com/a/apps/biomath/>) para que ocorresse o processo de desfosforilação, sendo a mistura submetido ao seguinte programa no termociclador: 1 hora a 37°C, 20 minutos a 65°C. Ao inserto não foi adicionado nada e a amostra foi colocada no termociclador por 20 minutos a 65°C. Após isso, vetor e inserto foram submetidos a uma reação de ligação para a produção do DNA recombinante de acordo com o protocolo de Sambrook e Russel (1989) utilizando a enzima “T4 DNA ligase” (New England Biolabs, Ipswich, Massachusetts, USA) na proporção de 3:1 (inserto:vetor).

4.6 Reação de Ligação do fragmento ao vetor de expressão

O Vetor pET28a (Novagen, Gibbstwon, New Jersey, EUA), conforme esquema apresentado na figura 3, foi digerido com as enzimas de restrição de modo a apresentar extremidades compatíveis com as do gene amplificado por PCR. Assim, vetor e inserto foram submetidos a uma reação de ligação para a

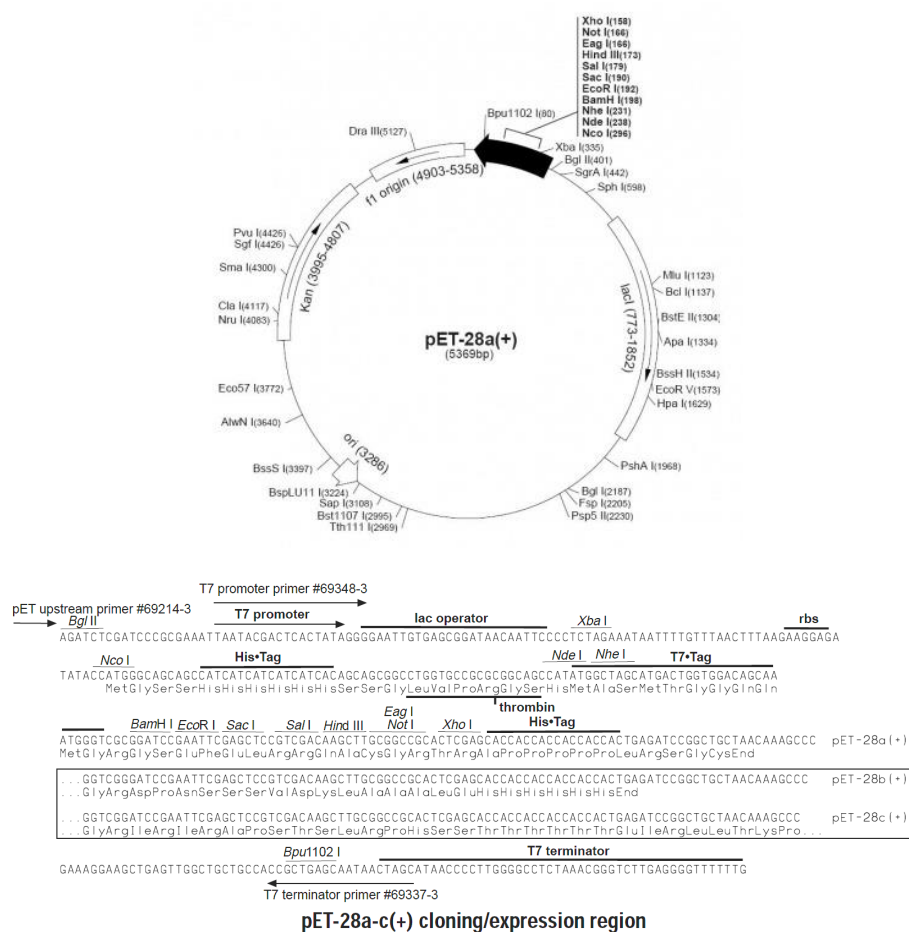
produção de DNA recombinante, realizada seguindo o protocolo de Sambrook e Russel (1989), onde utiliza T4 DNA ligase (New England Biolabs, Ipswich, Massachusetts, EUA) na proporção 3:1 (inserto:vetor).

Para a reação de ligação foi utilizada a seguinte fórmula:

$(\text{Vetor}(\text{ng}) \times \text{Inserto}(\text{kb})) / (\text{tamanho do vetor digerido} \times 3/1) = \text{quantidade em ng de inserto}$. Ou seja, $(55\text{ng} \times 2,340 \text{ kb}) / (5,369 \text{ kb} \times 3/1) = 71,9 \text{ ng}$ (quantidade de inserto necessária para a reação).

A reação de ligação se deu utilizando 2 μL de tampão T4 ligase 10x, 1,3 μL (72 ng) de inserto, 2,4 μL de vetor previamente digerido, 1,5 μL de T4 ligase e 12,8 μL de água, completando um volume de 20 μL . A reação foi realizada em aparelho termociclador por 16 horas a 16°C, e depois a amostra foi guardada em geladeira.

Figura 3 - Mapa do vetor pET-28a. Vetor usado para a expressão da ORF codificadora de enzima.



Fonte: <https://www.addgene.org/86463/>

4.7 Transformação bacteriana das células de *Escherichia coli*.

As transformações foram realizadas com células competentes de *E. coli* BL21, pRARE, Arctic e Origami previamente retiradas do freezer -80°C e descongeladas em banho de gelo por aproximadamente 10 min. Na reação foram utilizados 10 μL do DNA ligado (construção pET28a+ORF) e 200 μL da célula competente. A transformação foi feita por choque térmico e, para isso, a reação foi colocada por 10 min em banho de gelo e logo em seguida, submetida a 42°C , por 90 s, sendo recolocada imediatamente no banho de gelo por mais 2 min. Após a transformação, foram adicionados 970 μL de meio SOC [2% de triptona (m/v), 0,5% de extrato de levedura (m/v), 1 mL de NaCl 1 M, 1 mL de Mg^{2+} 1 M filtrado e 1 mL de glicose 1 M] para propiciar o desenvolvimento das células transformadas. As células foram submetidas a agitação de 230 rpm a 37°C por 90 min. Após incubação das células transformadas, alíquotas de 100 μL da cultura foram distribuídas em placas de petri contendo o meio LB sólido com 50 mg/mL de canamicina. As células transformadas foram incubadas a 37°C por 16 h.

A coleta dos clones foi realizada com palitos estéreis de madeira, selecionando um total de 10 colônias de forma aleatória. De uma mesma colônia com dois palitos diferentes, um dos palitos foi colocado em 3mL de meio LB com canamicina (50 $\mu\text{g}/\text{mL}$), o outro foi colocado em microtubo contendo 20 μL de tampão de lise e incubados no termociclador a 99°C por 30 min. Os tubos com LB + canamicina foram incubados sob agitação constante a 200 rpm, 37°C , por 24 h. Após este período, uma alíquota de 800 μL da cultura foi transferida para tubo contendo 200 μL de glicerol 100% esterilizado e estocados a -80°C . A confirmação da clonagem foi feita através da reação de PCR em colônias.

Os clones coletados foram depositados em microtubos de 0,25 mL estéreis e adicionados 20 μL de tampão de lise celular (KCl 50 mM, Tween 20 0,1% e Tris-HCl 10 mM, pH 8,3), sendo submetidos a 99°C por 30 min para a lise celular. Após, 4 μL do sobrenadante foram utilizados para a reação de PCR com os oligonucleotídeos iniciadores específicos para a ORF, e a visualização do resultado foi feita em eletroforese em gel de agarose a 1%.

4.8 Análise de sequenciamento da ORF

Após a PCR das colônias com os clones positivos foi realizada extração de DNA plasmidial utilizando o kit Wizard® Plus SV Minipreps DNA Purification System (Promega) para posterior sequenciamento e confirmação da inserção do fragmento no vetor pET28. Para a reação de sequenciamento os DNAs plasmidiais foram amplificados em microplacas nas seguintes condições: 100 ng de DNA plasmidial, 3,0 µL de tampão de sequenciamento 5X v3.1, 1,0 µL de BigDye v3.1 (Applied Biosystems Carlsbad, Califórnia, USA), 10 pmol do oligonucleotídeo, sendo que as reações foram feitas separadamente para *Forward* e *Reverse*, e água deionizada estéril para completar o volume de 10 µL. As placas foram levadas ao termociclador, com os seguintes ciclos: desnaturação a 96°C por 1 minuto; 39 ciclos de 96°C por 15 s; 65°C por 15 s, 60°C por 4 min e ciclo final a 4°C até serem retiradas.

Após a reação de sequenciamento procedeu-se à lavagem da placa, o DNA amplificado foi precipitado com 80 µL de isopropanol 75%, incubando-se em temperatura ambiente por 15 min, sendo posteriormente centrifugadas a 4.000 rpm por 45 min a 15°C, o sobrenadante foi descartado e as amostras foram lavadas duas vezes com 180 µL de etanol 70%, e centrifugadas a 4.000 rpm por 5 min. a 15°C. O sobrenadante foi descartado e o excesso de etanol foi retirado a vácuo por 5 min. As amostras foram ressuspendidas em 10 µL de formamida e incubadas por 5 min a 95°C, a fim de obter o DNA em fita simples e aplicadas no sequenciador automático ABI 3730 XL DNA Analyzer (Applied Biosystems, Foster City, CA) seguindo recomendações do fabricante. Essa etapa foi realizada pela técnica Dra. Camila Cesario Fernandes da *Facility* de Sequenciamento da UNESP de Jaboticabal (bit.ly/facility-fcav).

4.9 Construção do plasmídeo de expressão em *E. coli* (Fastbio)

O gene codificador de arabinofuranosidase também foi ligado ao vetor pET-28a(+) em um procedimento realizado pela empresa Fastbio. Para a construção do plasmídeo de expressão (pET-28a - CB10_2.116), para ligação, o vetor pET-28a(+) foi previamente digerido com as enzimas de restrição NdeI e XhoI. Essa construção

foi transformada em linhagens Origami 2(DE3) pLysS, Rosetta-gami 2(DE3) pLysS, Arctic e pRARE seguindo método de transformação por choque térmico.

4.10 Expressão da proteína recombinante

Para teste de expressão da proteínas de interesse, uma colônia isolada de *E. coli* transformada foi inoculada em 5 mL de meio de cultura LB com canamicina a 50 µg/mL, que permaneceu sob agitação durante 16 h, a 37°C, 200 r.p.m.. Alíquotas de 1 mL foram inoculadas em frascos contendo 50 mL de meio LB ou meio TB com canamicina, que permaneceram sob agitação de 200 rpm., 37°C, até atingir a fase logarítmica de crescimento, com D.O em 600 nm entre 0,8 e 1. A indução foi feita variando a concentração em 0,1 mM, 0,5 mM e 0,8 mM de IPTG à cultura, que permaneceu nas mesmas condições descritas por 16 h. Foram coletadas alíquotas de 1 mL da cultura no tempo zero (T0 - antes da indução) e no tempo 1 (T1 – 2 h após a indução), tempo 2 (T2 – 4 h após a indução), tempo 3 (T3 - 6 h após a indução), tempo 4 (T4 - 8 h após a indução) e tempo 5 (T5 - 16 h após a indução), essas alíquotas foram centrifugadas por 10 min a 1.700 xg, para que fosse possível analisar a expressão da proteína de interesse em comparação às proteínas totais. Diferentes condições foram testadas, com variações na temperatura, tempo de incubação, concentração de IPTG e rotação. Uma vez determinada a melhor condição de expressão de cada proteína, ensaios de expressão em volumes maiores foram realizados, seguindo o procedimento descrito anteriormente.

Para a extração das proteínas, as células foram coletadas por centrifugação a 12.000 xg por 20 min. A extração iniciou-se pela homogeneização do precipitado celular no tampão de extração (Tris-HCl 100 mM, pH 8, com NaCl 200 mM; Triton X100 0,2% e Imidazol 10 mM). As células foram lisadas por ultrassom em sonicador sendo 4 pulsos de 40 s com intervalos de 1 min. (para maiores volumes) ou com reagente BugBaster seguindo protocolo do fabricante. As amostras foram submetidas à centrifugação por 20 min, a 4°C, para separação de extratos solúveis (sobrenadante) e insolúveis (precipitado de restos celulares). Ambas as frações foram analisadas em gel de poliacrilamida (SDS-PAGE).

4.11 Eletroforese em gel de poliacrilamida (SDS-PAGE)

A análise da expressão das proteínas foi realizada através da eletroforese desnaturante em gel de poliacrilamida 10% SDS-PAGE. A corrida foi realizada a 150 V, com o tampão de corrida Tris-Glicina (50 mM de Tris-HCl; 150 mM de Glicina e 0,1% de SDS (m/v)). O preparo das amostras ocorreu com a adição do tampão de amostra 2X e fervura por 10 min a 95°C e o gel foi corado com Coomassie Brilliant Blue (40% de metanol (v/v); 10% de ácido acético glacial (v/v) e Coomassie R-250 a 0,1% (m/v)) e descorado em solução contendo 10% de ácido acético (v/v) e 10% de etanol (v/v). Cada etapa da obtenção da proteína foi analisada individualmente por meio de SDS-PAGE.

4.12 Recuperação da proteína em fração insolúvel

O *pellet* foi pesado após a lise celular (*pellet* do centrifugado para separar parte solúvel de insolúvel) e adicionou 3 vezes de solução 8 M de ureia. A amostra foi agitada em vortex para ressuspender o pellet na solução. Esse material foi deixado na bancada (à temperatura ambiente) por aproximadamente 30 min.

Na sequência a amostra foi centrifugada para clarificar o lisado, 15.000 xg por 20 min. Assim, se espera que a proteína esteja no sobrenadante, solubilizada em ureia. O sobrenadante foi transferido para um saco de diálise. As diálises foram feitas com no mínimo 2 h para cada troca de tampão. O experimento se deu da seguinte forma: de 8 para 4 M de ureia; depois de 4 para 2 M; depois de 2 para 1 M e por fim de 1 para 0 (diálise contra água). As amostras de cada troca de tampão foram coletadas para testar atividade, vale lembrar que às vezes não é possível retirar toda a ureia sem que a proteína precipite, também não se deve estocar enzimas/proteínas muito tempo em ureia, pois esta promove ligações entre proteínas.

4.13 Purificação por cromatografia de afinidade ao metal

Os extratos celulares obtidos na fração solúvel foram submetidos a uma cromatografia de afinidade a metal. A resina utilizada, HisTrap HP (GE Healthcare), é carregada com níquel, o que permite a purificação de proteínas com cauda de histidina (6 x His tag). Após lavagem da resina para retirar o etanol foi acrescentado o Tampão [10 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM de imidazol, 10% (v/v) glicerol] um gradiente de imidazol (de 20 mM a 1M), o qual também possui afinidade pelo níquel, compete com os aminoácidos da histidina pela ligação à resina e possibilita a eluição da proteína. As frações obtidas correspondentes aos picos foram coletadas e analisadas.

4.14 Ensaio de atividade enzimática

O ensaio enzimático foi realizado seguindo o protocolo adaptado proposto por Yang et al. (2015). O sistema de reação para a determinação da atividade de ABF consistiu em 25 µL de PNP-Abf (4 mM), 25 µL de tampão acetato pH 5,0, 20 µL de amostra de enzima serão incubados a 50 °C durante 1 hora, a reação foi parada por adição de 200 µL de Na₂CO₃ 4 M. A absorbância 410 nm foi lida utilizando um espectrofotômetro (Leitor de microplacas de absorvância - Spectramax plus 384).

5. RESULTADOS E DISCUSSÃO

5.1 Análises *in silico*

Inicialmente três ORFs correspondentes a arabinofuranosidases foram prospectadas do banco de dados (CB10) (<http://200.145.102.111/cb10/>) construído a partir do sequenciamento do consórcio microbiano, dando início aos estudos com a finalidade de selecionar uma sequência, sendo que um dos critérios a ser avaliado foi a presença de domínios conservados, uma vez que esses são fundamentais para o funcionamento adequado da enzima na fase de caracterização.

É sabido que enzimas microbianas que participam da degradação da parede celular da planta, as glicosil hidrolases (GHs), exibem uma variedade de estruturas modulares (YOSHIDA et al., 2010). Com isso, a identificação de domínios conservados que se caracterizam por pequenas regiões idênticas, em proteínas homólogas, permite que essas possam ser identificadas, caracterizadas e utilizadas para a localização de domínios similares em outras espécies. Esses domínios permanecem conservados, pois algumas regiões das proteínas são fundamentais para manter as suas propriedades biológicas, facilitando a sua localização por análises de alinhamentos de sequências (MALONE, 2006).

Os resultados da análise *in silico* da sequência CB10_2.116 confirmam o tamanho de 780 resíduos de aminoácidos identificada como uma α -L-arabinofuranosidase família de glicosil hidrolase GH 127 (<http://www.cazy.org/GH43.html>) com 80.3% de similaridade com uma proteína não caracterizada da bactéria *Chitinophaga jiangningensis*. A proteína é caracterizada por dois domínios conservados, um relacionado à família GH127 e um DUF (da abreviação inglesa para “domínio com função desconhecida”) conforme dados observados a partir de resultados de saída do banco Pfam (Figura 4).

A CB10_2.116 apresenta a homologia mais próxima de proteínas descritas como lignocelulolíticas de espécies de *Chitinophaga jiangningensis*, isso reflete a estreita relação entre os genes do gênero *Chitinophaga* e a capacidade de algumas de suas enzimas codificadas degradarem a biomassa hemicelulósica, sendo que ainda não foram catalogados na base de dados do CAZy, no caso de *Chitinophaga eiseniae* ao qual o DNA parcial foi comparado. A adição dessas proteínas a esse banco de dados é importante pois amplia a categorização de membros das famílias, adicionando informações sobre enzimas capazes de desconstruir a biomassa lignocelulósica.

Figura 4 - Coluna A: Apresenta o resultado de saída da plataforma Uniprot (<https://www.uniprot.org/>) para as sequências identificadas para arabinofuranosidases, onde a sequência CB10_282.6219 com 639 resíduos de aminoácidos e 79,7% de similaridade com uma Beta-xilosidase de *Chitinophaga jiangningensis*, seguida da ORF CB10_2116 780 aminoácidos e 80,3% de similaridade com uma proteína não caracterizada também de *Chitinophaga jiangningensis*, temos ainda a ORF CB10_61.2456 com 723 aminoácidos e 86,7% de similaridade com uma GH 43 de *Chitinophaga jiangningensis*. Coluna B: Apresenta dados de saída do SWISS-MODEL (<https://swissmodel.expasy.org/>) que é um servidor de modelagem de homologia de estrutura proteica, onde se é possível observar a localização dos domínios conservados na possível estrutura 3D da enzima. Coluna C: O servidor SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) prevê a presença e a localização dos sítios de clivagem do peptídeo sinal nas sequências de aminoácidos. O método incorpora uma previsão de locais de clivagem e uma predição de peptídeo sinal. Nas duas sequências que apresentaram peptídeo de sinal foi possível a remoção por não estar inserido em regiões de abrangência dos domínios da proteína.

CB10_282.6219

A

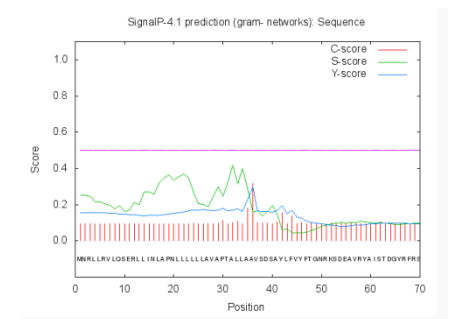
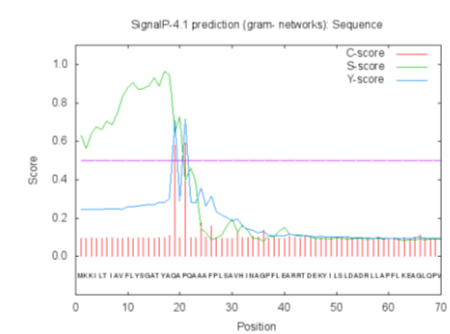
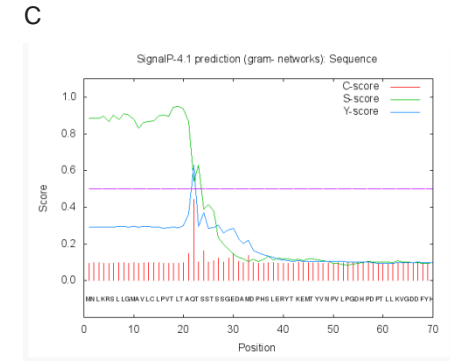
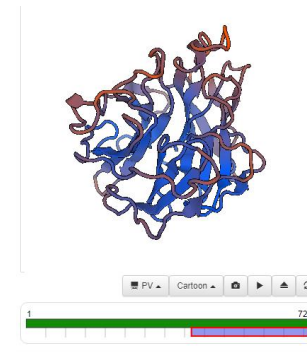
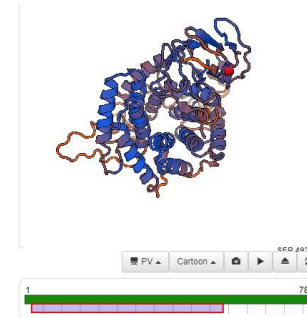
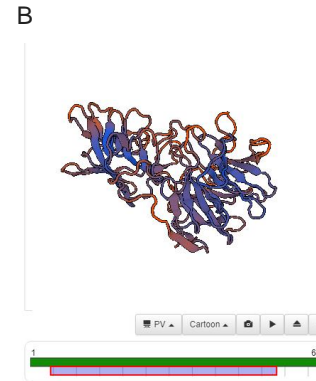
Entrada	Nomes de proteínas	Golpe de jogo	Identidade
A0A1M7CGS1	Beta- xilosidase (Chitinophaga Jiangningensis)		79,7%
D2QDL3	Familia de hidrolase de glicosídeos 43 (Spirosoma linguale (estirpe A...))		64,5%
GSTM12	Xilan 1,4-beta-xylosidase (Nastella koreensis (estirpe...))		66,1%
A0A1W6ES83	Glicósido hidrolase (Fibrella sp. ES10-3-2-2)		63,0%

CB10_2.116

Entry	Protein names	Match hit	Identity
A0A1M7LX72	Uncharacterized protein (Chitinophaga Jiangningensis)		80,3%
A0A1M7D3V2	Uncharacterized protein (Flavobacterium flevense)		51,5%
A0A1B9DMX2	Glycosyl hydrolase (Flavobacterium glycinis)		51,2%
A0A0Q0RP98	Acetyl-CoA carboxylase, biotin carboxylase (Flavobacterium daejeonense)		51,0%

CB10_61.2456

H1YA21	Glycoside hydrolase family 43 (Mucilaginibacter paludis DSM...)		61,8%
E4T0W0	Glycoside hydrolase family 43 (Paludibacter propionigenes...)		59,7%
A0A1S9PG71	Beta-xylosidase (Mucilaginibacter pedocola)		60,3%
A0A1M7CMK4	Glycosyl hydrolases family 43 (Chitinophaga Jiangningensis)		86,7%
A0A1MSW247	Glycosyl hydrolases family 43 (Wenyingshuangia marina)		59,1%
A0A1Q2HML1	Arabinoxylan arabinofuranohydrolase (Phycisphaerae bacterium L21...)		58,3%



Fonte: Autoria própria

As representações de modelagem por homologia (figura 4C) visam construir modelos 3D de estrutura da proteína usando estruturas experimentalmente determinadas de membros da família relacionados como modelos. O SWISS-MODEL é um servidor de modelagem de homologia de estrutura proteica totalmente automatizado, acessível através do servidor web ExPASy. A modelagem de homologia baseada na predição da estrutura da proteína é uma estratégia mais rápida de análise estrutural, comparada à difração de raio X ou . No entanto, uma limitação da modelagem de homologia é a disponibilidade de um modelo homólogo com uma identidade de sequência significativa (pelo menos maior do que 50%) e cobertura máxima de consulta no alinhamento (SHARMA; MEENA; KANWAR, 2018).

A visualização dos resultados de servidor SignalP 4.1 mostrou a presença de peptídeo sinal, porém esse não se encontrava inserido em nenhum dos domínios, possibilitando sua retirada no momento do desenho de primers. A observação e predição da estrutura se deu através de análise no programa PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) comprova essa hipótese conforme ilustra a figura 4C. Todas essas características fizeram da sequência CB10-2116 uma interessante candidata para a continuidade dos estudos.

A figura 5 apresenta a ilustração obtida a partir de análise em banco Pfam, onde se pode observar a presença de dois domínios, em verde um domínio relacionado a uma GH127 e em vermelho um domínio com função ainda desconhecida denominado de DUF. Também podemos observar o resultado da análise da estrutura através do programa PyMOL, onde foi possível visualizar a real posição do peptídeo sinal, e atrás disse considerar a remoção do mesmo da sequência inicial já que esse não se encontra inserido em nem um dos domínios a ser preservados.

Figura 5 – A: Ilustração obtida a partir de análise em banco Pfam. B: Análise da estrutura através do programa PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

A



B



Fonte: Autoria própria.

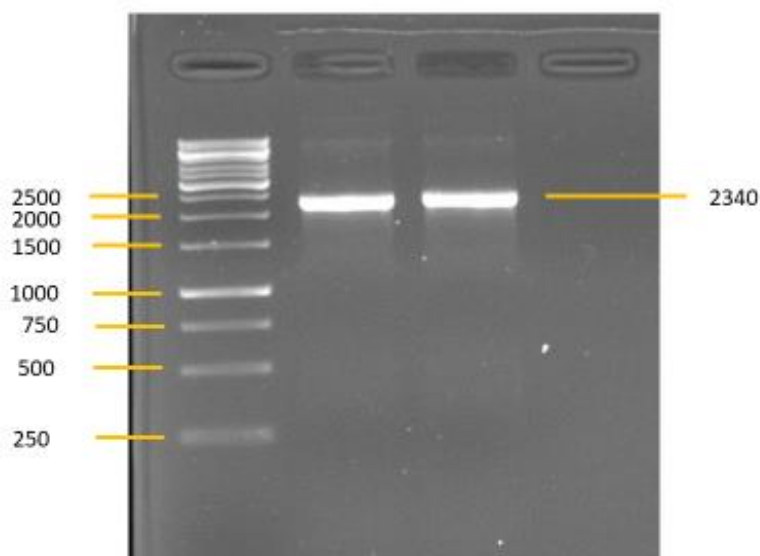
Segundo informações do Pfam (<http://pfam.xfam.org/family/PF16375>) as famílias "DUF" - DUF4986 (PF16375) presente na sequências em estudo - são anotadas com o domínio da função desconhecida (GOODACRE; GERLOFF; UETZ, 2014). Esta família, com cerca de 150 resíduos, localiza-se no C-terminal de algumas proteínas não caracterizadas em várias espécies de *Bacteroides* e *Bacillus*. A função desta família permanece desconhecida. Já a família GH127 foi descrita por Fujita et al. (2014) e contém atividade de β -L-arabinofuranosidase, que foi estabelecida para HypBA1 de *Bifidobacterium longum*. Os membros de GH127 eram conhecidos antigamente como membros da família Pfam DUF1680, que é

conservada em muitas espécies de bactérias, actinomicetos e plantas. O estabelecimento de GH127 por análise bioquímica resolveu o status de "domínio de função desconhecida" desta família.

5.2 Amplificação e clonagem dos genes no vetor pET28a

Para a confirmação da clonagem foi realizada a amplificação do gene por PCR (figura 6), as duas bandas que se destacam correspondem a mesma amostra feitas em microtubos. A amplificação do gene resultou em amplicons apresentando tamanho molecular compatível com o esperado, ou seja, fragmentos contendo 2.340 pares de base.

Figura 6 - Resultado da análise dos produtos de PCR em gel de agarose 1% utilizando Super Mix (Platinum Super Mix – High Fidelity - Invitrogen). Fragmento do estudo contendo 2.340 pares de base.



Fonte: Autoria própria.

A sequência codificadora de arabinofuranosidase identificadas nas etapas anteriores serviu como molde para a construção de oligonucleotídeos visando à amplificação dos genes para clonagem e expressão da proteína correspondente. A amplificação do gene codificado pela ORF CB10_2.116 a partir do DNA genômico de *Chitinophaga* sp resultou em fragmento do tamanho esperado, conforme evidenciado na figura 7. Uma vez obtido o fragmento de interesse, este foi

submetido a clivagem com as enzimas de restrição correspondente para clonagem no vetor de expressão pET28a.

Figura 7 - Gel de agarose da amplificação dos genes com oligonucleotídeos iniciadores específicos da PCR de colônias já clonadas confirmativa para os clones positivos. Sendo que os clones positivos podem ser visualizados nas posições 1,2,3,4,6,8 e 10.



Fonte: Autoria própria.

O produto digerido foi utilizado nas reações de ligação com o vetor para posterior transformação das células de *E. coli* BL21(DE3). As colônias foram analisadas por meio de PCR com oligonucleotídeos específicos para cada ORF. O perfil eletroforético em gel de agarose possibilitou verificar que dos 10 clones coletados da construção pET28 - CB-10 2.116, 7 clones tiveram amplificação da região de interesse.

Os clones positivos foram selecionados em placas com canamicina e verificados quanto ao fragmento esperado pela PCR de colônia, utilizando os *primers*. Todos os clones positivos foram sequenciados possibilitando a confirmação da identidade do gene clonado, o clone que apresentou melhores resultados diante da análise de sequenciamento foi escolhido para dar continuidade aos ensaios de expressão e caracterização enzimática (figura 8).

Figura 8 - Resultado do sequenciamento para *primer_F*, confirmando a inserção correta do fragmento.

CLUSTAL O(1.2.4) multiple sequence alignment

```

CB10_2.116      ----- 0
5_CB10_2.116_F.ab1  cgccagcaacagccaaccacaacaagccaggatcatcttccctctagaatattttgttt 60

CB10_2.116      ----- 0
5_CB10_2.116_F.ab1  aactttaagaaggagatataccatgggcagcagccatcatcatcatcacagcagcgg 120

CB10_2.116      ----- 0
5_CB10_2.116_F.ab1  cctggtgccgcgaggcagccatattggctagcatgactggtggacagcaaatgggtcgcgg 180

CB10_2.116      ----- 54
5_CB10_2.116_F.ab1  ----CCACAAGCAGCCGCCTTTCCCTTGAGCGCTGTACATAT-AAACGCGGGGCCGTTTC 240
                    atcctccacaagcagccgccttcccttgagcgtgtacatagagaacgcggggccgttc
                    ***** .*****

CB10_2.116      CTGGAGGCGCGCCGCACGGATGAAAAGTACATCCTTTCCCTTGACGCGGACAGGCTACTG 114
5_CB10_2.116_F.ab1  ctggaggcgcgcccgcagggatgaaaagtacatccttcccttgacgaggacaggctactg 300
                    *****

CB10_2.116      GCCCCCTTCTGAAGGAAGCGGGCTGCAGCCTGTAAAAGATAGTTACGGCAACTGGGAA 174
5_CB10_2.116_F.ab1  gcccccttctgaaggaagcgggcctgcagcctgtaaaagatagttacggcaactgggaa 360
                    *****

CB10_2.116      AAAGACGGGCTGAACGGCCACATCGGGGGCCATTATTTATCCGCGCTTGCCTGATGTAT 234
5_CB10_2.116_F.ab1  aaagacgggctgaacggccacatcggggccattatTTATCCGCGCTTGCCTGATGTAT 420
                    *****

CB10_2.116      GCCAGCACCGCAATGAAGTGTACCATCAACGACTGCAATACATGCTGGACCAGCTGGAA 294
5_CB10_2.116_F.ab1  gccagcaccggcaatgaagtgtaccatcaacgactgcaatacatgctggaccagctggaa 480
                    *****

CB10_2.116      CGTTGCCAGCAGGCCAACAAAAACGGCTATATCGGCGGTGTCCCAAACGGCAAGGCCGTT 354
5_CB10_2.116_F.ab1  cgttgccagcaggccaacaaaaacggctatATCGGCGGTGTCCCAAACGGCAAGGCCGTT 540
                    *****

CB10_2.116      TGGGACGAAGTGGC-TGCCGGCAA-TGTCAGCGGGTAAAAAGCGCTGGGTGCCCTGGT 412
5_CB10_2.116_F.ab1  agggacaacatggcttgcggcaatggtcagcgggtgaaaaaccctgggcgccttgct 600

```

Fonte: Autoria própria

5.3 Ensaio de expressão e extração das proteínas recombinantes

As análises iniciais das propriedades de cada proteína, tais como massa molecular e ponto isoelétrico (pI) teóricos, foram feitas por meio da ferramenta ProtParam (Tabela 2). Os dados obtidos com estas análises foram utilizados em ensaios de expressão para a identificação das proteínas de interesse no gel de poliacrilamida.

Tabela 2 - Análise no protparam para a sequência codificadora de arabinofuranosidases. (<https://web.expasy.org/cgi-bin/protparam/protparam>).

	Número de Resíduos	Massa molecular (kDa)	pI estimado
CB10_2.116	780	87,99	8,58

Fonte: Autoria própria

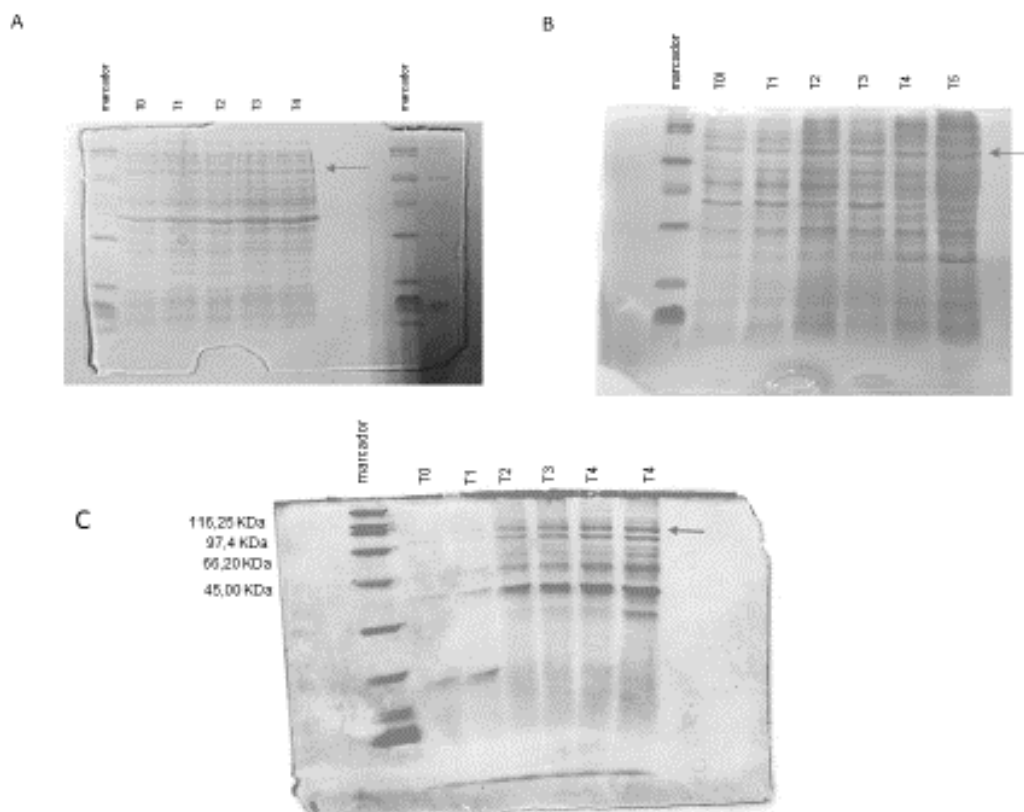
Conforme descrevem Rosano e Ceccarelli (2014) se buscarmos na literatura por uma cepa de *E. coli* adequada para uso como hospedeiro, produzirá dezenas de possíveis candidatas. Todas elas têm vantagens e desvantagens. No entanto, algo a ter em mente é que muitas são especiais e são usadas em situações específicas. Algumas características genéticas de BL21 fazem dela uma das mais utilizadas; essas células são deficientes na protease Lon, que degrada muitas proteínas exógenas. Outro gene ausente do genoma dos ancestrais da BL21 é o que codifica a protease da membrana externa OmpT, cuja função é degradar as proteínas extracelulares.

As células transformadas em linhagem de *E. coli* BL21 apresentam uma banda compatível com a banda de interesse, e aparentemente a variação na concentração do indutor IPTG (20°C, T4 em 0.8mM) melhorou a expressão da banda correspondente (figura 9), sendo que os tempos de maior exposição ao indutor provocaram a maior intensidade da banda conforme ilustra a figura 9C.

Camargo et al. (2018) também conseguiu melhores índices de expressão com maior tempo de indução; em seu trabalho apresenta resultados de expressão

de uma arabinofuranosidase de *Clostridium thermocellum* em recombinante de BL21 (DE3) onde foi utilizado 0,5 mM de IPTG e incubação por 20 h a 20 °C.

Figura 9 - Análise eletroforética da fração solúvel do lisado de células transformadas em linhagem BL21 nas seguintes condições de indução: A- Indução (25° C - 0,1 mM IPTG – meio LB); B- Indução (20° C - 0,5 mM IPTG - Meio TB); C- Indução (20°C - 0,8 mM de IPTG – Meio LB). Onde, T0 (sem indução), T1 (após 2 horas de indução), T2 (após 4 horas de indução), T3 (após 6 horas de indução), T4 (após 8 horas de indução), T5 (após 16 horas de indução). A seta indica a banda compatível com a massa molecular (87,99 kDa) da enzima recombinante.



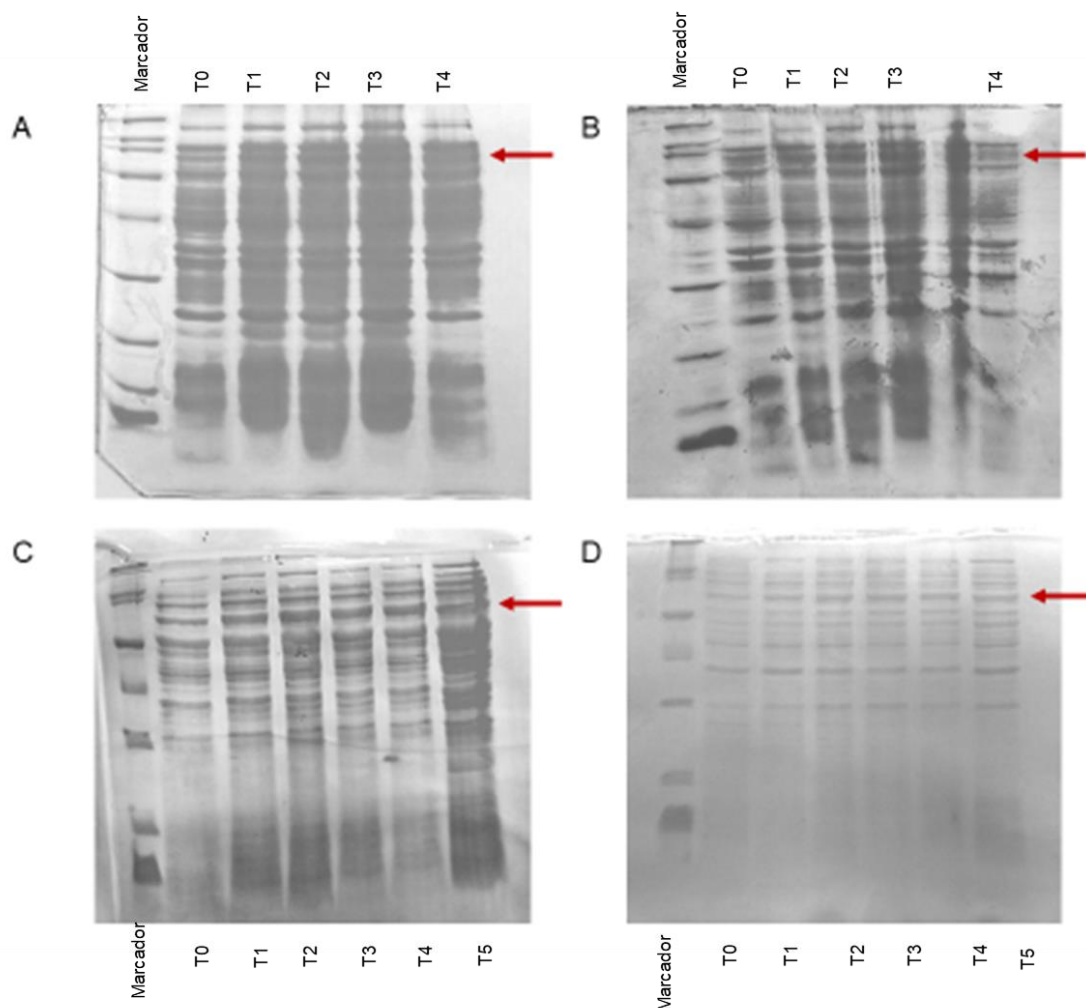
Fonte: Autoria própria.

Quanto à transformação em diferentes linhagens de *E.coli* (figura 10) podemos notar a presença da banda, considerando que a indução em pRARE e Arctic (20°C - 0,5 mM de IPTG – Meio LB) foram as que apresentaram uma maior coerência, a expressão não aumenta ao longo do tempo e no caso de indução em Arctic o T2 (após 4 horas de indução) se mostra pouco mais intenso que os demais tempos.

Camargo et al. (2018) também explicam que quando a expressão da proteína recombinante é baixa e não pode ser aumentada pelos mecanismos propostos, então o rendimento volumétrico da proteína desejada pode ser aumentado pelo crescimento da cultura para atingir densidades mais elevadas, que no caso desse estudo foram feitos testes com DO entre 0,8 e 1. Isto pode ser conseguido alterando alguns parâmetros, como a composição do meio e proporcionando melhor aeração por agitação vigorosa.

Um problema recorrente na expressão de proteínas está relacionado à formação de corpúsculos de inclusão. Para sanar esse problema a síntese de proteína recombinante deve ser realizada na faixa de 15 a 25°C (ROSANO; CECCARELLI, 2014), embora um relato tenha descrito a expressão bem sucedida a 4°C por 72 h (SAN-MIGUEL; PÉREZ-BERMÚDEZ; GAVIDIA, 2013).

Figura 10 – Fração solúvel do lisado de células transformadas em diferentes linhagem de *E. coli* nas seguintes condições de indução: A- Roseta Indução (18° C - 0,5 mM IPTG – meio LB); B- Origami Indução (18° C - 0,5 mM IPTG - Meio TB); C- Arctic Indução (20°C - 0,5 mM de IPTG – Meio LB). D- pRARE Indução (20°C - 0,5 mM de IPTG – Meio LB). As colunas nos géis apresentam a seguinte ordem: Marcador, T0 (sem indução), T1 (após 2 horas de indução), T2 (após 4 horas de indução), T3 (após 6 horas de indução), T4 (após 8 horas de indução), T5 (após 16 horas de indução).



Fonte: Autoria própria.

No entanto, quando se trabalha na extremidade inferior da faixa de temperatura, o crescimento é mais lento e as taxas reduzidas de síntese podem resultar em menores rendimentos de proteína. Além disso, o dobramento de proteínas pode ser afetado, já que a rede de chaperonas pode não ser tão eficiente. A estirpe *ArticExpress* possui a chaperonina adaptada ao frio e a co-chaperonina

da bactéria psicrófila *Oleispira antarctica*. As chaperoninas exibem altas atividades de redobragem a temperaturas de 4–12 °C e conferem uma capacidade aumentada para *E. coli* crescer a temperaturas mais baixas (ROSANO; CECCARELLI, 2014).

5.4 Purificação e atividade enzimática

Os testes de purificação e caracterização não atingiram resultados favoráveis, sendo que por diversas vezes se tentou a purificação usando colunas HisTrap HP que por sua vez são embaladas com resina de afinidade Ni Sepharose High Performance (HP). Esta resina consiste em esferas de agarose altamente reticuladas com um grupo quelante acoplado. O grupo quelante é pré-carregado com níquel, que retém seletivamente proteínas com grupos histidina expostos.

Quanto aos testes de atividade enzimática, não foi possível, em nenhuma das tentativas detectar atividade para arabinofuranosidase.

6. CONSIDERAÇÕES FINAIS

Neste estudo foi possível identificar e analisar *in silico* três ORFs que codificam arabinofuranosidases no banco de dados CB10 disponibilizado pelo LBMP – UNESP de Jaboticabal, das quais uma (a CB10_2.116) foi selecionada para clonagem, expressão e caracterização.

A clonagem do fragmento foi realizada com sucesso em vetor de expressão pET28a. Foram realizados testes de expressão da proteína, e conforme se pode visualizar nos géis apresentados se pode dizer que a expressão foi atingida. A banda correspondente ao tamanho da proteína de interesse que apresenta 87,99 kDa aparece nos géis de poliacrilamida, porém não foi possível notar a superexpressão da enzima; testes foram realizados modificando variáveis que pudessem influenciar na expressão enzimática.

Sendo assim, não foi possível alcançar a expressão ideal dessa enzima, em virtude do sistema escolhido. Acreditamos que um novo sistema de expressão, por exemplo em *Pichia pastoris*, possa ser testado para se obter a proteína purificada, ou testes que utilizem diferentes substratos enzimáticos.

REFERÊNCIAS

BRIAND, L, et al. Heterologous expression and characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Clostridium thermocellum* B8. **Enzyme and Microbial Technology**, v. 109, p. 74–83, 2018.

BRIENZO, M.; CARVALHO, W.; MILAGRES, A.M. Xylooligosaccharides production from alkali-pretreated sugarcane bagasse using xylanases from *Thermoascus aurantiacus*. **Applied Biochemistry and Biotechnology**, v. 162, n. 4, p. 1195-1205, 2010.

CAMARGO, B. et al. Heterologous expression and characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Clostridium thermocellum* B8. **Enzyme and Microbial Technology**, v. 109, p. 74–83, 2018.

CARVALHO, D. et al. A halotolerant bifunctional β -xylosidase/ α -L-arabinofuranosidase from *Colletotrichum graminicola*: Purification and biochemical characterization. **International Journal Of Biological Macromolecules**, v. 114, p.741-750, 2018.

CARVALHO, N. L.; BORTOLINI, J. G.; BARCELLOS, A. L. Biocombustíveis: uma opção para o desenvolvimento sustentável. **Revista Gestão e Desenvolvimento em Contexto- Gedecon**, v. 1, n. 1, p.32-50, jan. 2015.

CINTRA, L C *et al.* Characterization of a recombinant xylose tolerant β -xylosidase from *Humicola grisea* var. *thermoidea* and its use in sugarcane bagasse hydrolysis. **International Journal Of Biological Macromolecules**, v. 105, p.262-271, dez. 2017.

CHAPLA, D.; PANDIT, P.; SHAH, A. Production of xylooligosaccharides from corncob xylan by fungal xylanase and their utilization by probiotics. **Bioresource Technology**, v. 115, p. 215-221, 2012.

CHEN, C.; HUANG, H.; WU, C. H. Protein Bioinformatics Databases and Resources. **Protein Bioinformatics**, v. 1, n. 1, p.3-39, 2017.

COLLINS, T. GERDAY, C. FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **FEMS Microbiology Reviews**, v. 29, n.1, p 3–23, 2005.

CONTESINI, F. J. et al. Structural and functional characterization of a highly secreted α -L-arabinofuranosidase (GH62) from *Aspergillus nidulans* grown on sugarcane bagasse. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1865, n. 12, p. 1758–1769, 2017.

ESPINDOLA, F.S. et al. Bioinformatic resources applied on the omic sciences as genomic, transcriptomic, proteomic, interatomic and metabolomic. **Biosci. J**, v. 26, n. 3, p 463-477, 2010.

FINN, R D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research**, v. 44, n. 1, p.279-285, 15 dez. 2016.

FUJITA, K. et al. Characterization of a Novel β -L-Arabinofuranosidase in *Bifidobacterium longum*. **Journal of Biological Chemistry**, v. 286, n. 44, p. 38079–38085, 2014.

GONÇALVES, T. A. et al. Functional characterization and synergic action of fungal xylanase and arabinofuranosidase for production of xylooligosaccharides. **Bioresource Technology**, v. 119, p.293-299, 2015.

GOODACRE, N. F.; GERLOFF, D. L.; UETZ, P. Protein Domains of Unknown Function Are Essential in Bacteria. **mBio**, v. 5, n. 1, p. e00744-13, 2014.

JIA, L. et al. Synergistic degradation of arabinoxylan by free and immobilized xylanases and arabinofuranosidase. **Biochemical Engineering Journal**, v. 114, p. 268–275, 2015.

KAUR, KUMAR, KAUR. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. **International Journal of Biological Macromolecules**, v. 106, p. 803-822, 2018.

KISHI, L. T. et al. Draft Genome Sequence of a *Chitinophaga* Strain Isolated from a Lignocellulose Biomass-Degrading Consortium. **Genome Announcements**, v. 5, n. 3, p. e01056-16, jan. 2017.

KURNIATI, A.; DARMOKOESOEMO, H.; PUSPANINGSIH, N. N. T. Scanning Electron Microscope Analysis of Rice Straw Degradation by a Treatment with α -L-arabinofuranosidase. **Procedia Chemistry**, v. 18, p. 63–68, 2016.

LOMBARD, V. et al. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research**, v. 42, n. 1, p.490-495, 2013.

MALONE, G. et al. Prospecção de Genes em Bibliotecas de cDNA. **Revista Brasileira Agrocência**, v. 12, n. 1, p. 7–13, 2006.

MECHELKE, M. et al. Characterization of the arabinoxylan-degrading machinery of the thermophilic bacterium *Herbinix hemicellulosilytica*: Six new xylanases, three arabinofuranosidases and one xylosidase. **Journal Of Biotechnology**, v. 257, p.122-130, 2017.

MENEZES, C. R.; BARRETO, A. R. Biodegradação de resíduos lignocelulósicos por fungos basidiomicetos: Caracterização dos resíduos e estudo do complexo enzimático fúngico. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v. 19, n. 2, p.1365-1391, ago. 2015.

MOHSEN DEHNAVI, S.; PAZUKI, G.; VOSSOUGH, M. PEGylated silica-enzyme nanoconjugates: A new frontier in large scale separation of α -amylase. **Scientific Reports**, v. 5, n. December, p. 1–9, 2015.

NUMAN, M. T.; BHOSLE, N. B. α -L-Arabinofuranosidases: the potential applications in biotechnology. **Journal Of Industrial Microbiology & Biotechnology**, v. 33, n. 4, p.247-260, 30 dez. 2005.

PETERSEN, T. N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v. 8, n. 10, p.785-786, 29 set. 2011.

ROSANO, G. L.; CECCARELLI, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. **Frontiers In Microbiology**, v. 5, n. 1, p.1-17, 17 abr. 2014.

ROZENDO, A. S. et al. Strategies for induction in the *Leishmania chagasi* recombinant antigens production. **Revista Saúde e Ciência**, v. 3, n. 3, p.174-188, dez. 2014.

SAMBROOK, J.; RUSSELL, D. W. Molecular cloning: a laboratory manual. 3. ed. New York: Cold Spring: Cold Spring Harbor, 1989.

SAN-MIGUEL, T; PÉREZ-BERMÚDEZ, P.; GAVIDIA, I. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. **Licensee Springer**, v. 2, n. 89. 2013.

SANTOS, F.; OSVALDO A.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p.253-259, mar. 2003.

SANTOS, F. A. et al. Potencial da palha de cana-de-açúcar para produção de etanol. **Química Nova**, v. 35, n. 5, p.1004-1010, 2012.

SAXENA et al. The Basic Concepts of Molecular Modeling. **Methods in Enzymology**, v. 467, p 307-334, 2009.

SHALLOM, D. et al. The identification of the acid-base catalyst of alphaarabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. **FEBS Letters**, v. 514, n. 2-3, p. 163-167, 2002.

SHARMA, A; MEENA, K. R; KANWAR, S. S. Molecular characterization and bioinformatics studies of a lipase from *Bacillus thermoamylovorans* BHK67. **International Journal Of Biological Macromolecules**, v. 107, p.2131-2140, 2018.

SINGHA, T. K. et al. Efficient genetic approaches for improvement of plasmid based expression of recombinant protein in *Escherichia coli*: A review. **Process Biochemistry**, India, v. 55, n. 1, p.17-31, 2017.

SOUZA, L.L; RHODEN, S. A; PHAMPHILE, J.A. The importance of omics as a tool for the study of exploration of microorganisms: prospects and challenges. **Revista UNINGÁ**, v. 18, n. 2, p. 16-21. 2014.

WILKENS. et al. GH62 arabinofuranosidases: Structure, function and applications. **Biotechnology Advances**, v. 35, n. 6, p.792-804, 2017.

YANG, W. et al. A novel bifunctional GH51 exo- α -L-arabinofuranosidase/endo-xylanase from *Alicyclobacillus* sp. A4 with significant biomass-degrading capacity. **Biotechnology for Biofuels**, v. 8, n.1, p. 197, 2015.

YASIR, M. et al. *Chitinophaga eiseniae* sp. nov., isolated from vermicompost. **Int J Syst Evol Microbiol** v. 61, p. 2373–2378, 2011.

YOSHIDA, S. et al. Domain Analysis of a Modular-L-Arabinofuranosidase with a Unique Carbohydrate Binding Strategy from the Fiber-Degrading Bacterium *Fibrobacter succinogenes* S85. **Journal Of Bacteriology**, v. 192, n. 20, p.5424-5436, 2010.

TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, 30/07/2019

Cintia Lionela A. Menezes

Assinatura do autor